Population Structure of the *Bacillus cereus* group

Margaret Barker

A thesis submitted for the degree of

Doctor of Philosophy

School of Life Sciences

Heriot-Watt University

Edinburgh, UK

April 2006

ABSTRACT

The *Bacillus cereus* group of bacteria comprises *B. anthracis, B. cereus, B. mycoides, B. pseudomycoides, B. thuringiensis* and *B. weihenstephanensis*. Species status has been allocated to these taxa largely according to pathogenic properties. *B. anthracis* is the causative agent of anthrax in ungulates and humans. *B. thuringiensis* is primarily an insect pathogen and *B. cereus* is associated with food poisoning and occasionally soft tissue infections in humans.

One hundred and forty-six strains of the *B. cereus* group were examined by multilocus sequence typing (MLST) in which partial sequences for seven housekeeping genes (*glpF, gmk, ilvD, pta, pur, pycA* and *tpi*) were generated to provide a definitive sequence type (ST) for each strain. Statistical analyses of the data using pairwise comparisons between groups for (i) Fst (gene flow), (ii) shared mutations and (iii) fixed differences confirmed that the present designation of separate species status for members of the *B. cereus* group was inappropriate. Comparison of neighbour joining (NJ) trees derived from the concatenated sequence data with trees constructed for each allele individually indicated limited recombination between strains and a largely clonal structure to the group. Three major clades were recovered: clade 1 was made up of *B. anthracis, B. cereus* and rare *B. thuringiensis* strains; clade 2 comprised a heterogeneous mixture of *B. thuringiensis* and *B. cereus* strains while clade 3 was composed of strains of *B. cereus, B. mycoides* and *B. weihenstephanensis.* Two *B. pseudomycoides* strains were distant outliers from the main tree. Four lineages were recognised in both clades 1 and 2 based on shared mutations within the lineages and fixed differences between them.

B. anthracis strains and the emetic toxin-producing strains of *B. cereus* formed two clones within clade 1. A clonal group of entomopathogenic *B. thuringiensis* strains was identified in clade 2 and named the ‘Sotto’ lineage (after the predicted founder group). Strains of *B. cereus* that had been isolated from human wound infections and septicaemia, on the other hand, were distributed over clades 1 and 2, and were not restricted to a particular clonal group. Similarly, some serotypes of *B. thuringiensis* were found to have a clonal structure while others were heterogeneous. Representative strains from several serotypes of *B. thuringiensis* were examined by the RAPD (random

amplified polymorphic DNA) method. Serovars israelensis and thuringiensis were strongly clonal, morrisoni and tolworthi were partially clonal while darmstadiensis and canadensis were heterogeneous. Serotype, MLST profile and RAPD did not always correlate with delta-endotoxin *cry* gene content. This may be due to the *cry* genes being located on plasmids and subject to transfer between strains.

MLST does not support the separate species status of *B. anthracis, B. cereus, B. mycoides, B. pseudomycoides, B. thuringiensis* and *B. weihenstephanensis* and an alternative classification based on DNA sequence data is proposed based on three main clades with nine distinct lineages. The proposed lineages were named to be consistent with current nomenclature, as far as possible.

Keywords: - *Bacillus cereus* group, multilocus sequence typing (MLST), sequence type (ST), clade, lineage, recombination, clonal structure, RAPD, *cry* genes.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Fergus Priest, for making it possible for me to carry out interesting research work in his group and to study part-time for a Ph.D. I have appreciated his knowledge, enthusiasm and helpful advice throughout this project.

Thanks are also due to all of my colleagues in the laboratory over the years, particularly to Dr Elise Cachat, Dr Sylvie van Beek, Dr Adelaida de Gavira, Dr Yang Yu, Dr Sam de Costa and Dr Jen Brown.

I gratefully acknowledge the help and advice on DNA sequencing techniques and data analysis given to me by Dr Martin Maiden and the other members of his research group at Oxford University. In particular, I am grateful to Dr Keith Jolley for his work in setting up the on-line MLST database for the *B. cereus* group.

I also acknowledge the help and support I have received from many members of the academic and technical staff in the School of Life Sciences at Heriot-Watt University.

Finally, thanks and appreciation is due to my husband, Derek for his continued support and encouragement at home and also to my son, Edward and daughter, Lucy.

ACADEMIC REGISTRY
Research Thesis Submission

Name*:*	Margaret BARKER		
School/PGI:	School of Life Sciences		
Version: *(i.e. First, Resubmission, Final)*	Final	Degree Sought:	Ph. D

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

1) the thesis embodies the results of my own work and has been composed by myself
2) where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
3) the thesis is the correct version of the thesis for submission*.
4) my thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying, subject to such conditions as the Librarian may require
5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

* *Please note that it is the responsibility of the candidate to ensure that the correct version of the thesis is submitted.*

Signature of Candidate*:*		Date:	

Submission

Submitted By *(name in capitals):*	Margaret BARKER
Signature of Individual Submitting:	
Date Submitted:	

For Completion in Academic Registry

Received in the Academic Registry by *(name in capitals):*			
Method of Submission *(Handed in to Academic Registry; posted through internal/external mail):*			
Signature:		Date:	

DECLARATION

I, Margaret Barker, hereby declare that I am the author of this thesis. All the work described in this thesis is my own except where stated in the text. The work presented here has not been accepted in any previous application for a higher degree. All the sources of information have been consulted by myself and are acknowledged by means of reference.

(Margaret Barker)

CONTENTS

Chapter 1 INTRODUCTION

Chapter 3 RESULTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY

Abbreviation	
APS	Ammonium persulphate
BGSC	Bacillus Genetic Stock Centre
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate Generic name for dATP, dCTP, dGTP and dTTP.
EDTA	Ethylenediaminetetracetic acid
EM	Electron microscope
Fst	'F statistic' (from the program DnaSP) measures the extent of DNA divergence between two populations and computes the level of gene flow.
h	Hours
GLB	Gel loading buffer
m	minutes
MEE	Multilocus enzyme electrophoresis
Mol%	Molar percentage
MLST	Multilocus sequence typing
NA	Nutrient agar
NB	Nutrient broth
NSM	Nutrient agar and salts medium
NYSM	Nutrient agar, yeast extract and salts medium
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonyl fluoride
s	Seconds
RAPD	Random amplified polymorphic DNA
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
TBE	Tris borate EDTA
TBS	Tris-buffered saline
TEMED	N,N,N',N'-Tetramethylethylenediamine
TSA	Tryptone soya agar
TSB	Tryptone soya broth

TTBS	Tween and tris-buffered saline
v/v	Volume for volume
w/v	Weight for volume

Solutions

BCIP 5-bromo-4-chloro-indoyl phosphate (for western blot)

BCIP Na salt	20 mg ml^{-1}
Water to 1 ml	

Colour development buffer (for western blot)

Tris base	12.11 g
Sodium chloride	5.84 g
HCl	As required to adjust pH to 9.0
water to 1 litre	

Concentration is 0.1 M tris and 0.1 M NaCl, pH is 9.0

Colour Development solution (for Western Blot)

NBT solution	45 µl
BCIP solution	87.5 µl
Colour development buffer	10 ml

GLB Gel Loading Buffer (for DNA electrophoresis)

glycerol	25 % v/v
SDS	1 % v/v
bromophenol blue	0.25 % w/v (omit for RAPDs)
xylene cyanol	0.25 % w/v (omit for RAPDs)
EDTA pH 8	150 mM

2 µl GLB + 10 µl sample loaded into well

NBT Nitro blue tetrazolium (for Western Blot)

dimethylformamide	70 %
NBT	75 mg ml^{-1}

Resolving gel buffer (4x) (for SDS-Page)

Tris base (1.5 M)	18.17g
20% SDS	2 ml
H Cl to pH 8.8 and water to 100 ml	

Resolving gel (9%) (for SDS-Page)

Acrylamide (40%)	8.0 ml
Bisacrylamide (2%)	5.5 ml
Buffer (4x resolving)	9.25 ml
water	12.3 ml
10 % APS	0.2 ml
TEMED	20 µl

Running buffer (10 x) (for SDS-Page)

Tris base	30.28 g
Glycine	144.25 g
SDS	10 g
Water to 1 litre	

Sample buffer (2x) (for SDS-Page)

Tris-HCl 0.5 M pH 6.8	6.25 ml
SDS	2 g
glycerol	9 ml
2-mercaptoethanol	5 ml
Bromophenol blue (0.1 %)	1 ml
Water to 50 ml	

Sequencing Buffer (5x) (for DNA sequencing)

Tris – HCl	400 mM
$MgCl_2$	10 mM
pH 9	

Stacking gel buffer (4x) (for SDS Page)

Tris base (0.5 M)	6.06g
20 % SDS	2 ml

H Cl to pH 6.8 and water to 100 ml

Stacking gel (4.5 %) (for SDS-Page)

Acrylamide (40%)	1.1 ml
Bisacrylamide	0.7 ml
Buffer (4x stacking gel)	2.5 ml
water	5.6 ml
10 % APS	30 µl
TEMED	10 µl

Stain (for SDS-Page)

Coomassie blue (R250)	0.5 g
methanol	250 ml
Glacial acetic acid	50 ml
water	200 ml

TAE Tris acetate Buffer (50x) (for gel electrophoresis)

Tris base	242 g
Glacial acetic acid	57.1 ml
0.5 M EDTA (pH 8)	100 ml

Water to 1 litre

Working solution is 0.04 M tris-acetate and 0.001 M EDTA.

TAQ mix (5x) (for PCR)

Taq NH_4 buffer (10x)	500 µl
dATP (100 mM)	10 µl
dCTP (100 mM)	10 µl
dGTP (100 mM)	10 µl
dTTP (100 mM)	10 µl
sterile water	460 µl

TBE Tris borate EDTA (5x) (for gel electrophoresis)

Tris base	54 g
Boric acid	27.5 g
0.5 M EDTA (pH 8)	20 ml

Working solution is 0.089 M tris-borate and 0.002 M EDTA

TBS Tris buffered saline (for Western Blot)

Tris base	3.03 g
Sodium chloride	29.22 g

Water to 1 litre

Final concentration is 20 mM tris and 500 mM NaCl, pH is 7.5.

Transfer Buffer (for Western Blot)

Tris base	3.03 g
glycine	14.4 g

Water to 1 litre

Final concentration is 25 mM tris and 192 mM glycine, pH is 8.3.

TTBS (tris buffered saline with tween)

As for TBS but with the addition of 0.05 % tween-20

PUBLICATIONS

Publications arising from this research: -

Barker, M., Thakker, B., Priest, F.G. 2005. Multilocus sequence typing reveals that *Bacillus cereus* strains isolated from clinical infections have distinct phylogenetic origins. FEMS Microbiol. Letters. **245:** 179-184.

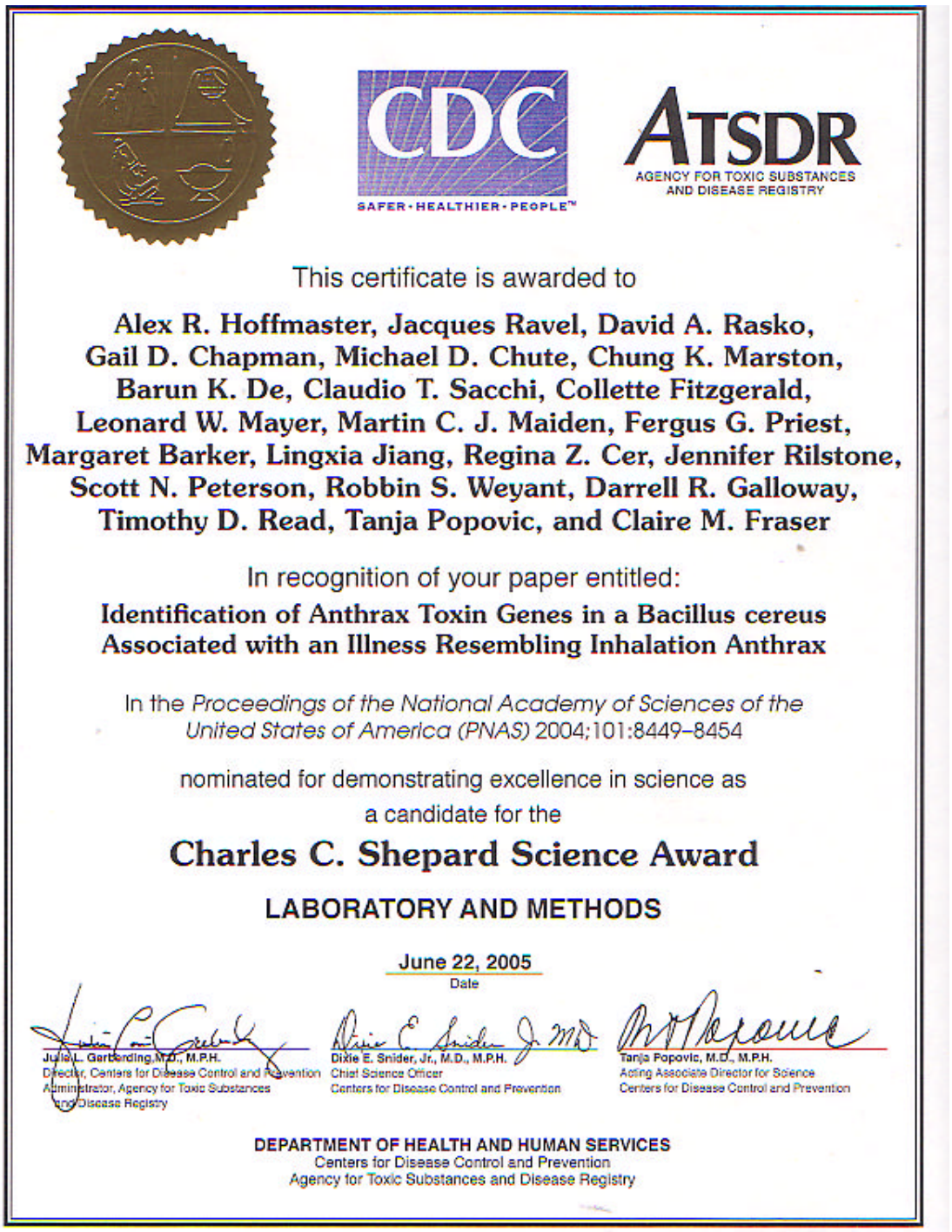
Priest, F.G., Barker, M., Baillie, L.W.J., Holmes, E.C., Maiden, M.C.J. 2004. Population Structure and Evolution of the *Bacillus cereus* Group. J. Bacteriol. **186:** 7959-7970.

(The two papers listed above are included at the end of this thesis).

Publication associated with this research (MLST database): -

*** Hoffmaster, A.R., Ravel, J., Rasko, D.A., Chapman, G.D., Chute, M.D., Marston, C.K., De, B.K., Sacchi, C.T., Fitzgerald, C., Mayer, L.W., Maiden, M.C.J., Priest, F.G., Barker, M., Jiang, L., Cer, R.Z., Rilstone, J., Peterson, S.N., Weyant, R.S., Galloway, D.R., Read, T.D., Popovic, T., Fraser, C.M.** 2004. Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. PNAS. **101:** 8449-8454.

- This publication was awarded the Charles Shepard science award for demonstrating excellence in science (reproduced on page xxv).



CDC
SAFER·HEALTHIER·PEOPLE™

ATSDR
AGENCY FOR TOXIC SUBSTANCES AND DISEASE REGISTRY

This certificate is awarded to

Alex R. Hoffmaster, Jacques Ravel, David A. Rasko, Gail D. Chapman, Michael D. Chute, Chung K. Marston, Barun K. De, Claudio T. Sacchi, Collette Fitzgerald, Leonard W. Mayer, Martin C. J. Maiden, Fergus G. Priest, Margaret Barker, Lingxia Jiang, Regina Z. Cer, Jennifer Rilstone, Scott N. Peterson, Robbin S. Weyant, Darrell R. Galloway, Timothy D. Read, Tanja Popovic, and Claire M. Fraser

In recognition of your paper entitled:

Identification of Anthrax Toxin Genes in a Bacillus cereus Associated with an Illness Resembling Inhalation Anthrax

In the *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* 2004;101:8449–8454

nominated for demonstrating excellence in science as a candidate for the

Charles C. Shepard Science Award

LABORATORY AND METHODS

June 22, 2005
Date

Julie L. Gerberding, M.D., M.P.H.
Director, Centers for Disease Control and Prevention
Administrator, Agency for Toxic Substances and Disease Registry

Dixie E. Snider, Jr., M.D., M.P.H.
Chief Science Officer
Centers for Disease Control and Prevention

Tanja Popovic, M.D., M.P.H.
Acting Associate Director for Science
Centers for Disease Control and Prevention

DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
Agency for Toxic Substances and Disease Registry

Chapter 1 INTRODUCTION

This project examines the diversity found within the *Bacillus cereus* group of bacteria. Species included in this group range from the highly pathogenic *B. anthracis* to harmless *B. mycoides* soil isolates. It also encompasses food poisoning strains of *B. cereus* and the insect pathogen *B. thuringiensis* (characterized by the production of a crystal protein toxin). The taxonomic structure of the group is unclear and the separate species status of some members of the *B. cereus* group of bacteria is questionable.

> **'But it's simple. *Bacillus anthracis* causes anthrax, *Bacillus thuringiensis* kills insects and *Bacillus cereus* is……..…...…the other one'.**

The above quotation was made by my supervisor, Professor F. G. Priest, at the 5th Anthrax Conference in Nice, 2003, during a discussion in which some strains had been described as *B. thuringiensis* on the basis of flagellar 'H' serotyping, but did not have an insecticidal crystal protein.

In this project, I clarify the phylogenetic structure of the *B. cereus* group of bacteria and have developed a generally accessible database, with the facility for additional strains to be added. As the database grows, the structure of the group will become more accurately defined and the separate species status of some sub-groups can be re-examined. Additionally, the relative position of any new strains can be plotted on a phylogenetic tree and this may give an indication about its phylogenetic origin and possible pathogenic potential.

1.1 Taxonomy

The *B. cereus* group of bacteria is included in the phylum *Firmicutes* (bacteria with a thick cell wall) and belongs to the family *Bacillaceae.*

1.1.1 The family *Bacillaceae*

The family *Bacillaceae* includes Gram-positive, usually rod-shaped (one genus is spherical) bacteria that produce distinctive endospores. The mol % G + C is generally low (35 – 55%). Most are motile but can be aerobic, facultative or anaerobic (Sneath, 1986). There are five genera: -

Genus I	*Bacillus*	Aerobic or facultative anaerobic, usually produce catalase.
Genus II	*Sporolactobacillus*	Microaerophilic, do not produce catalase.
Genus III	*Clostridium*	Anaerobic, sulphate not reduced to sulphide.
Genus IV	*Desulfotomaculum*	Anaerobic, sulphate reduced to sulphide.
Genus V	*Sporosarcina*	Spherical cells, in packets.

1.1.2 The genus *Bacillus*

In 1872, Ferdinand Cohn described rod-shaped bacteria with a filamentous growth as '*Bacillus'*. A few years later, it was noticed that some species of *Bacillus* were able to produce spores and by 1920 the genus *Bacillus* was re-defined to include only spore-forming rods which could grow aerobically. In 1973, 'The Judicial Commission of the International Committee on Systematic Bacteriology' reviewed the names given to bacteria and retained only those which were well described. Additionally, each named species must have a defined 'type strain' deposited in a recognized culture collection. In 1980, Skerman *et al.* compiled the named bacteria in 'List of Bacterial Names with Standing in Nomenclature'. The list includes, alphabetically and chronologically, the nomenclature of prokaryotes and the nomenclatural changes as cited in the 'Approved Lists of Bacterial Names' or validly published in the 'International Journal of Systematic Bacteriology' (IJSB) or in the 'International Journal of Systematic and Evolutionary Microbiology' (IJSEM). About 180 *Bacillus* species are now listed (May, 2005) and details are available on-line at http://ijs.sgmjournals.org.

There is a problem with the definition of *Bacillus* as Gram-positive, spore-forming bacteria, in that older cultures often appear to be Gram-negative, and some strains do not readily produce spores. Such factors may lead to a strain being wrongly identified (Harwood *et al.*, 1989). The genus *Bacillus* is very diverse and suggestions to sub-

divide the genus have been made, initially on the basis of numerical taxonomy studies (Logan & Berkeley, 1981; Priest *et al.*, 1988) and later, from rRNA and rRNA gene sequence analyses (Ash et al., 1991; Joung & Cote, 2002; Blackwood *et al.,* 2004). Indeed, over the past decade the *Bacillus sensu lato* has been re-defined and the BGSC catalog of strains now lists 15 genera (June 2005). Many endospore-forming bacteria have been recently discovered in extreme environments, e.g. halophiles, acidophiles, alkaliphiles and thermophiles. Small subunit rRNA gene sequencing has identified these bacteria as close relatives of *Bacillus.*

1.2 The *Bacillus cereus* group

The *B. cereus* group is distinct from but closely related to *B. subtilis*, the type species of the genus *Bacillus* (Figure 1.1). Strains from the *B. cereus* group of bacteria are characteristically gram-positive, spore forming, motile, facultatively anaerobic rods. When grown on Nutrient Agar, a characteristic feature for *B. cereus, B. anthracis, B. mycoides* and *B. thuringiensis* is a cell width of >1 μm. Cell length is more variable and depends on growth conditions (Harwood, 1989).

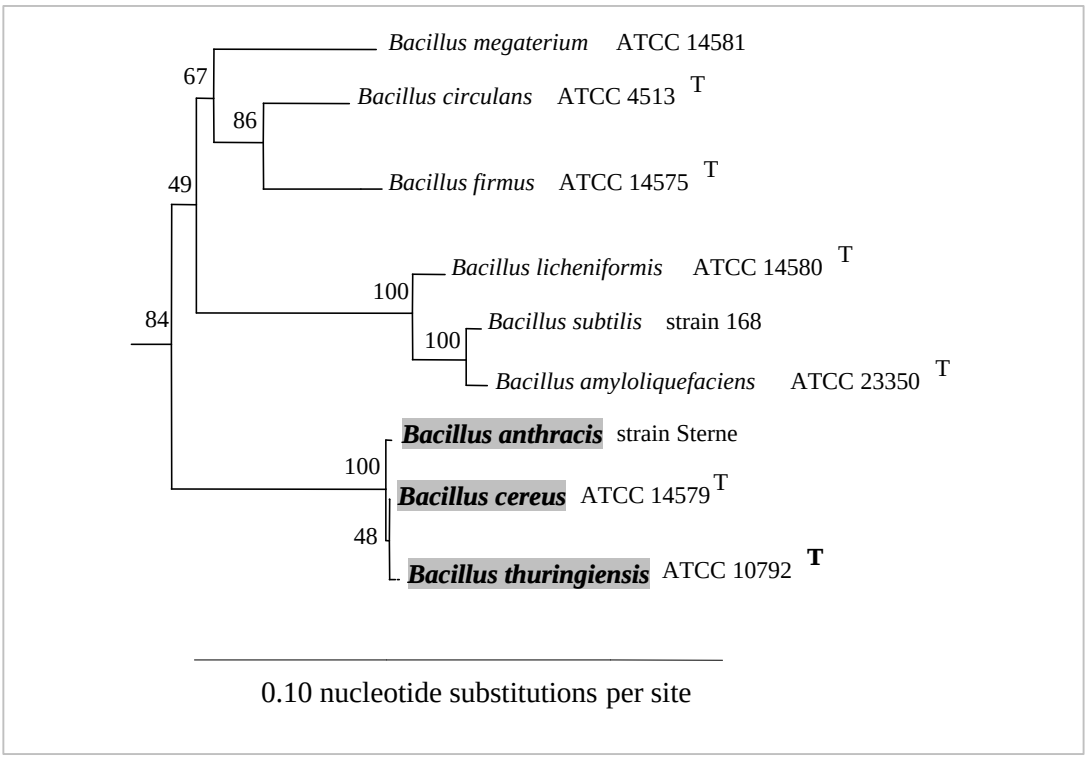


Figure 1.1 Phylogenetic tree showing the position of the *B. cereus* group (shaded) in relation to *B. subtilis*, derived from an alignment of 16S rRNA gene sequences. (Reproduced from Priest & Dewar, 2000).

Strains have been isolated from many sources, including soil, insect larvae, wound infections and a variety of foods, especially rice and milk products (Gordon *et al.*, 1974; Granum & Lund. 1997; Priest & Dewar, 2000; Priest *et al.*, 2004). The group includes some important pathogens of mammals and insects. Some strains of the *B. cereus* group are important pathogens: - *B. anthracis* causes the disease anthrax in humans and cattle. The genes causing the disease are located on two plasmids, pX01 and pX02 (Okinaka *et al.*, 1999). *B. thuringiensis* produces a crystal protein which is toxic to insects. The genes for the insecticidal protein are generally plasmid encoded (Schnepf *et al.*, 1998). Some strains of *B. cereus, B. thuringiensis, B. mycoides, B. pseudomycoides, B. weihanstephansis* and *B. thuringiensis* have toxin genes which, if expressed, may cause outbreaks of food poisoning and these toxin genes are chromosomal (Pruß *et al.*, 1999).

Currently, the *B. cereus* group comprises several separate species – *B. anthracis, B. cereus, B. mycoides, B. pseudomycoides, B. thuringiensis* and *B. weihenstephanensis.* For many years, there has been controversy concerning the separate species status given to these bacteria (Helgason *et al.*, 2000 b).

Although the type strain of *B. cereus*, ATCC 14579, dates from 1887 (Frankland and Frankland) *B. cereus* was not recognized as a separate species from *B. subtilis* until 1936, when the International Congress for Microbiology defined *B. subtilis* more strictly. Subsequently, *B. cereus* was considered to be the 'parent species' and *anthracis, thuringiensis* and *mycoides* were defined as varieties (Gordon *et al*., 1974) but this classification was not generally accepted because of the pathogenic attributes of *B. anthracis* and *B. thuringiensis* strains. Separation of these taxa has been based on phenotypic characteristics, including pathogenic host range, despite their acknowledged genetic closeness. The key features which have been used to separate the species, e.g. the virulence of *B. anthracis* and the insecticidal crystal toxins of *B. thuringiensis* are located on plasmids, which can be lost or transferred to another strain (Helgason *et al.*, 2000 b; Rasko *et al*., 2005). Despite these problems separate species status was allocated to *B. anthracis, B. cereus, B. thuringiensis B. mycoides, B. pseudomycoides* and *B. weihenstephanensis* in the 'List of Bacterial Names with Standing in Nomenclature' which is available on-line at http://ijs.sgmjournals.org. Some phenotypic differences between species of the *B. cereus* group are shown in Table 1.1.

Table 1.1 Distinguishing features of closely related species within the *B. cereus* group.

Species	Colony growth	Motility	Haemo lysis	Penicillin sensitivity	Para sporal crystal inclusion
B. anthracis	normal	-	-	+	-
B. cereus	normal	+	+	-	-
B. mycoides	rhizoidal	-	+/-	-	-
B. pseudomycoides	rhizoidal	-	-	-	-
B. thuringiensis	normal	+	+	-	+
B. weihenstephanensis	normal	-	+/-	-	-

1.2.1 *B. anthracis*

B. anthracis is phenotypically different from *B. cereus* in that *B. anthracis* strains are penicillin-sensitive, non-haemolytic, non-motile and encapsulated (Table 1.1).

Anthrax is one of the earliest recorded diseases of humans and other mammals. The oldest records relate to Biblical descriptions of the fifth and sixth Egyptian plagues at the time of Moses (ca. 1250 B.C.) (Turnbull, 2002). In 1857, David Livingstone wrote about a disease in Africa called 'horse-sickness'. '*…..great numbers of zebras are found dead with masses of foam at the nostrils, exactly as occurs in the common horse-sickness. The production of the malignant carbuncle…….by the flesh when eaten, is another proof of the disease of the tame and wild being identical'*

In 1877, Robert Koch cultured *B. anthracis*, noted that it produced endospores, and injected the bacterium into animals, which then developed anthrax. This was the first proof that bacteria can cause disease. Spores can survive for many years in dry soil but do not germinate in soil or puddles, requiring the growth conditions found in a mammalian host. When the animal dies, bacteria from the carcass sporulate, completing the cycle of infection. Tabaniid flies, blowflies and mosquitos have been shown to spread *B. anthracis* after feeding on infected animals. Insects appear to have a symbiotic relationship with *B. anthracis* and act as a vector for the organism (Jensen *et al.*, 2003).

The whole genome of *B. anthracis* Ames-strain has been sequenced and is about 5.23 megabases. (Read *et al.*, 2003). The genes responsible for the pathogenicity of *B. anthracis* are located on two plasmids and both plasmids must be present for virulence

to be expressed (Okinaka *et al.,* 1999). Most of the putative chromosomal proteins coding for virulence in *B. anthracis* have homologues in *B. cereus* and *B. thuringiensis*. PlcR, the gene regulator which controls the expression of many of the virulence genes in *B. cereus* is truncated in *B. anthracis*, due to a mutation rendering it phenotypically distinct (Slamti *et al.,* 2004).

1.2.2 *B. cereus*

B. cereus strains are widespread in the environment and the majority do not appear to be pathogenic. However, some strains are associated with diarrhoeal or emetic food poisoning outbreaks (Granum, 1994; Granum & Lund, 1997, Pirttijarvi *et al.,* 1998, 1999, 2000), while others have been isolated from eye infections (Drobniewski 1993), periodontitis (Helgason *et al.,* 2000 a) and burn wounds (Damgaard *et al.,* 1997). Other strains have been recovered from the soil (Helgason *et al.,* 1998; Vilas-Boas *et al.,* 2002) and as symbionts in insect intestines. In some arthropods a filamentous form of *B. cereus* has been found colonizing the gut (the Arthromitus stage). When repeatedly cultured and transferred, the filaments become shorter and motility declines. Eventually the bacteria grow mainly as single cells but with some diplocells (Margulis *et al.,* 1998). Colonization of mosquito larvae, other insects and earthworms has also been observed (Jensen *et al.,* 2003).

Comparison of the whole genome sequences for *B. anthracis* A2012 and *B. cereus* ATCC 14579 found that 75-80% of the genes were conserved and probably indicate a common ancestor for the two species. Studies of the potential properties of the common genes indicated that proteins, peptides and amino acids were probably the main nutrient sources, suggesting that the ancestral bacterium may have inhabited insect intestines rather than soil. Genes encoding nearly all of the toxins found in known pathogenic strains of *B. cereus* were identified in the non-pathogenic *B. cereus* ATCC 14579 (Ivanova *et al.,* 2003).

1.2.3 *B. mycoides*

B. mycoides is characterized by spreading rhizoid growth on agar plates. The earliest reference to *B. mycoides* is from a publication by Flügge in 1886. While recognizing the distinct growth pattern, Gordon *et al.,* (1973) described *B. mycoides* as a variety of *B. cereus* because she and her colleagues noted that the rhizoid growth could be lost on

sub-culture and the resultant strains were indistinguishable from *B. cereus*. Of thirteen strains of rhizoidal *B. mycoides* examined by Gordon, twelve were non-motile, indicating that there is not a link between the invasive growth habit and motility.

Later, Nakamura & Jackson (1995) clarified the status of *B. mycoides* by DNA relatedness and fatty acid composition of 58 strains from the *B. cereus* group, including 32 strains morphologically identified as *B. mycoides.* Nakamura & Jackson found that *B. cereus* (16 strains) and *B. thuringiensis* (11 strains) were 'distinct, but genetically moderately closely related species', with 59 to 69 % DNA similarity. *B. mycoides* was more distantly related, sharing 22 to 44 % DNA similarity with *B. cereus*. Moreover, the *B. mycoides* strains separated into two distinct hybridization groups with about 30% DNA reassociation. One they assigned as *B. mycoides* and the second was later named *B. pseudomycoides.*

1.2.4 *B. pseudomycoides*

In 1998, Nakamura officially proposed the recognition of *B. pseudomycoides* as a separate species (Nakamura, 1998). Like *B. mycoides, B. pseudomycoides* strains form rhizoidal colonies which spread to cover the culture plate. 16S rRNA sequences and physiological criteria could not distinguish between *B. mycoides* and *B. pseudomycoides.* The distinguishing features were differences in fatty acid composition and DNA hybridization of about 30%.

1.2.5 *B. thuringiensis*

B. thuringiensis was isolated in 1901 in Japan from diseased silkworms and was called '*Bacillus sotto'*. In 1911, the bacterium was re-discovered (independently) as a pathogen of flour moths in the province of Thuringia in Germany by Berliner and named *B. thuringiensis* (reviewed by Priest & Dewar, 2000). *B. thuringiensis* is well known as an insect pathogen and is distinguished from *B. cereus* by the presence of various protein parasporal crystals or delta-endotoxins. Indeed, if the crystal proteins are discounted, *B. cereus* and *B. thuringiensis* appear to be phenotypically identical (Priest & Dewar, 2000). Strains of *B. thuringiensis* may have different combinations of delta-endotoxins and may be active against one or more of the insect orders *Lepidoptera*, *Diptera* and *Coleoptera*. There are reports that some *B. thuringiensis* strains may be

toxic to insects from other orders and also to nematodes, mites and protozoa (Schnepf *et al.*, 1998).

Commercial preparations based on mixtures of dried spores and crystals are the most prevalent biological insecticide and account for about 2% of the world-wide agrochemical market. Toxin genes from *B. thuringiensis* have been genetically engineered into various crop plants to protect them from insect attack (Schnepf *et al.*, 1998). The sub-specific classification of *B. thuringiensis* is described in more detail in section 1.4.

1.2.6 *B. weihenstephanensis*

B. weihenstephanensis was described as a new species by Lechner *et al.* (1998). The main feature of *B. weihenstephanensis* strains is cold tolerance (growth at 7ºC or less). The authors also reported the presence of a 16S rRNA signature sequence and also the presence of a variant cold shock gene, *cspA.*

1.3 Toxins of the *B. cereus* group

Members of the *B. cereus* group produce a wide range of toxins. Some of the genes encoding the toxins are located on plasmids, others are genomic and one, the emetic toxin, cereulide is produced by non-ribosomal peptide synthetase (NRPS). Fortunately, not all strains are able to produce all toxins and most strains have evolved into specialists, probably to exploit a particular environmental niche.

1.3.1 *B. anthracis* toxins

The genes responsible for *B. anthracis* virulence are located on two plasmids, which have been sequenced (Okinaka *et al.*, 1999). They are pX01 (181 kb) that encodes the lethal toxin complex and pX02 (93.5 kb), which contains the genes responsible for the poly-?-D-glutamic acid capsule. Recent work has focused on the mechanism of the anthrax toxin complex. Anthrax toxin is in two parts - a receptor-binding part, the protective antigen (PA), and two catalytic factors – edema factor (EF) and lethal factor (LF). PA is able to form pores in the lipid bilayers of cell membranes, probably by changing the pH to acidic. The mechanism of interaction between EF and LF alters the

conformation of PA and allows the EF/LF complex to cross into vesicles inside endosomes. The toxins are transported along the vesicles to target areas of the cytoplasm near the cell nucleus where they interfere with phosphorelay pathways, reducing kinase activity. The mechanisms of interaction between EF and LF and their effects on the immune system are still being studied (Abrami *et al.,* 2005; Scobie & Young, 2005).

Most of the putative chromosomal proteins coding for virulence in *B. anthracis* have homologues in *B. cereus* and *B. thuringiensis*. PlcR, the gene regulator which controls the expression of many of the virulence genes in *B. cereus* is truncated in *B. anthracis* (Slamti *et al.,* 2004).

Recently, an unusual strain of *B. cereus*, G9241, was isolated from a patient with pneumonia (Hoffmaster *et al.,* 2004). This *B. cereus* strain was found to have the anthrax toxin genes encoded on a plasmid similar to pX01. However, pX02, the plasmid containing the poly-?-D-glutamic acid capsule genes was not present. Instead, another plasmid, pBC218 was identified and found to encode a polysaccharide capsule cluster. This provided the bacterium with an atypical capsulated phenotype enabling it to invade the lungs and cause a disease resembling inhalation anthrax (Hoffmaster *et al.,* 2004).

1.3.2 *B. cereus* group toxins (excluding *B. anthracis*)

Although primarily associated with food poisoning, strains from the *B. cereus* group (excluding *B. anthracis* and *B. cereus* G9241) are also capable of causing various local and systemic infections (Turnbull *et al.,* 1979; Helgason *et al.,* 2000a; Kotiranta *et al.,* 2000; Barker *et al.,* 2005). Some of the infections attributed to *B. cereus* are shown in Figure 1.2. Additionally, Turnbull pointed out that *B. cereus* strains are often found in hospital infections, but are not thought to be the cause of the disease and are regarded as contaminants. The infections are opportunistic and healthy individuals usually recover quickly, however, there is a more serious problem with immunologically compromised patients, the old and the very young (Drobniewski, 1993).

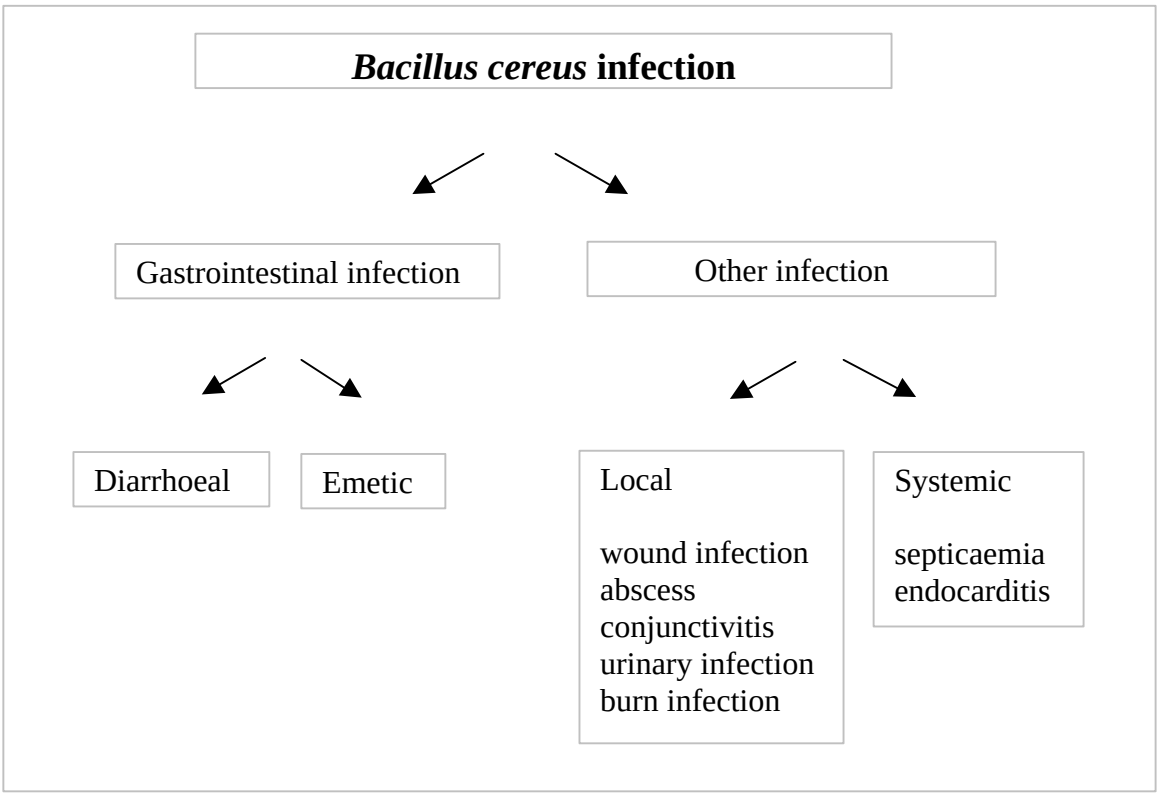


Figure 1.2 *Bacillus cereus* infections (adapted from Drobniewski *et al.*, 1993).

The emetic-toxin producing *B. cereus* strains have been implicated in outbreaks of food poisoning usually associated with cooked rice dishes or pasta (Kotiranta *et al.,* 2000; Agata *et al.,* 2002). Other strains from the *B. cereus* group cause diarrhoea and are often associated with milk, milk products, meat and vegetables (Beecher & Macmillan, 1990; Drobniewski, 1993). In addition to the food poisoning toxins, *B. cereus* is known to produce at least two haemolysins and three phospholipases C. One of the phospholipases, sphingomyelinase is also a haemolysin. Most of these virulence factors are under the control of a gene regulator, PlcR (Agaisse *et al.,* 1999; Beecher & Wong, 2000).

1.3.2.i Emetic toxin

The emetic toxin produces nausea and vomiting about 1 to 5 hours after eating contaminated food and the acute effects last for 6 to 24 hours. In 95% of emetic *B. cereus* food poisoning cases, fried or cooked rice has been involved (Horwood *et al.,* 2004). The emetic toxin is only produced by a small group of *B. cereus* strains (Agata *et al.,* 1995, 1996, 2002). The toxin is a dodecadepsipeptide (cereulide) that is capable of forming vacuoles in cells. The effect of the toxin was tested on an animal model, *Suncus murinus* (the house shrew), and episodes of vomiting at different concentrations of toxin

were recorded: - 'The average frequency of vomiting was 6 times and these episodes occurred in a few minutes' (Agata *et al.,* 1996). The antibiotic valinomycin is structurally similar to cereulide and was also found to induce vomiting, but at much higher concentrations. It was noted later that none of the strains associated with emetic toxin synthesis were able to hydrolyze starch or produced the diarrhoeal toxin. The vacuoles that are formed in cells exposed to cereulide are actually swollen mitochondria and later a bioassay was developed using mitochondrial swelling and loss of motility in boar spermatozoa (Andersson *et al.,* 1998). Cereulide is produced by a non-ribosomal peptide synthetase (NRPS). Other substances produced by NRPS genes include surfactin, gramicidin and ß-lactam antibiotics (Horwood *et al.,* 2004). Conserved sequences were found in the NRPS genes, and PCR primers have now been designed to detect the NRPS genes of emetic strains of *B. cereus*.

A study using ribotyping found that the emetic strains formed a clonal group (Pirttijarvi *et al.,* 1999), despite having different countries of origin and isolation dates. A recent polyphasic study confirmed the distinctly clonal character of the emetic toxin producing strains of *B. cereus.* A collection of 90 *B. cereus* strains was studied using RAPD and M13 PCR typing, MLST, 16S and 23S rRNA gene and spacer region sequencing, Fourier transform infrared spectroscopy, exoprotein profiling, PCR for the detection of enterotoxin genes, starch hydrolysis and haemolysis. The four phylogenetic trees produced from these studies recovered 3 to 4 groups, with the emetic strains clustering together in all cases. The emetic type of *B. cereus* was further characterized by an inability to hydrolyse starch, little or no haemolytic activity and absence of the genes coding for the haemolytic enterotoxin *hbl* (Ehling-Schulz *et al.,* 2005).

1.3.2.ii Enterotoxins

Food poisoning symptoms arising from the ingestion of *B. cereus* cells or spores include diarrhoea and abdominal pain after an incubation period of 8 to16 hours, and last for 12 to 24 hours (Granum, 1994). *B. cereus* cells produce enterotoxin during their exponential growth phase, while growing at 37 °C in the small intestine. Vomiting or fever is not usually associated with this type of food poisoning, although a person could be infected with an emetic toxin producing strain and a diarrhoeal strain at the same time.

The production of at least two different enterotoxins, causing diarrhoea, has been associated with bacteria from the *B. cereus* group (Granum, 1994; Granum & Lund, 1997; Kotiranta *et al.,* 2000). The two enterotoxins are Nhe, a three-component non-haemolytic enterotoxin (Granum *et al.,* 1999; Lund & Granum, 1999) and Hbl, a three-component haemolysin (Heinrichs *et al.,* 1993. Ryan *et al.,*1997; Prüß *et al.,* 1999). A necrotic cytotoxin, CytK has also been reported (Lund, *et al.,* 2000). The non-haemolytic enterotoxin, Nhe is produced by nearly all *B. cereus* strains, while Hbl is present in about half of the strains tested (Lindbäck *et al.,* 2004)

Other virulence factors such as phospholipases C and sphingomyelinase have been described. Several virulence factors may be expressed simultaneously and some have been found to operate synergistically (Gilmore *et al.,*1989; Beecher & Wong, 2000; Pomerantsev *et al.,* 2003). Many of the virulence genes are controlled by *plcR* (a pleiotropic gene regulator) (Lereclus *et al.,*1996; Agaisse *et al.,*1999).

1.3.2.iii Non-haemolytic toxin Nhe

The Nhe enterotoxin was isolated from a strain of *B. cereus* implicated in a food poisoning outbreak in Norway in 1988, and reported to be similar to Hbl, but non-haemolytic (Granum & Lund, 1997; Granum *et al.,* 1999). The operon was found to consist of three genes, *nheA, nheB* and *nheC* (Figure 1.3). Transcription of the genes is under the control of PlcR. In 1999, Lund and Granum sequenced and purified *nheC,* a novel gene encoding a 105-kDa protein of Nhe, and found that the amino acid sequence had a high similarity to collagenases from *Clostridium histolyticum* and *Clostridium perfringens* (Lund & Granum, 1999). Western immunoblotting showed that NheB was the only component that bound to Vero cells (African green monkey kidney cells). Cytolytic activity in the Vero cell assay required that the concentration of NheC must be about eight times lower than the concentrations of NheA and NheB. NheC was not detectable in 2D gel electrophoresis and this may be because it would only be present in very small quantities (Gohar *et al.,* 2002; Lindbäck *et al.,* 2004).

Nhe operon

Hbl operon

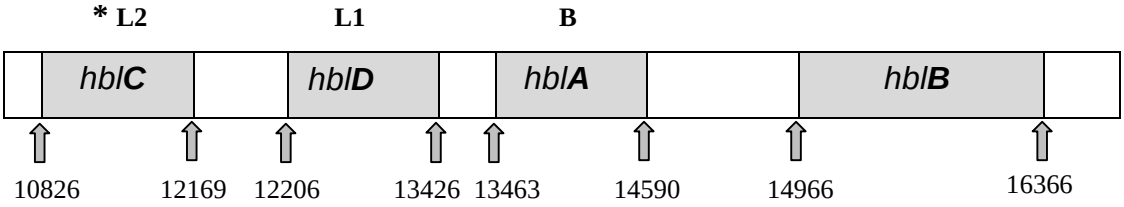


Figure 1.3 Diagrammatic representation of the nhe and hbl operons (not to scale).

* Hbl operon – 'L2' is the lytic component 2, 'L1' is the lytic component 1 and 'B' is the binding component.

1.3.2.iv Haemolytic toxin (Hbl)

A *B. cereus* diarrhoeal enterotoxin was found to have haemolytic activity and to be made up of three components, (B, L1 and L2) of which B binds to erythrocytes before the L components act to lyse the cells (Beecher & Macmillan, 1990; 1991). The genes encoding the B, L1 and L2 components have been cloned and characterized. The L1 (1154 bases) and L2 (1343 bases) components are separated by only 37 bases. The three genes are arranged in tandem and are co-transcribed, making this unit an operon. The genes are in the order 'L2 (*hblC*) - L1(*hblD*) – B(*hblA*)', as shown in Figure 1.3. The three components operate together to produce a ringed zone of haemolysis on blood agar plates and have also been shown to cause fluid accumulation in rabbit ileal loops (the standard test for diarrhoeal enterotoxins) (Heinrichs *et al.,* 1993; Ryan *et al.,* 1997).

1.3.2.v Similarities between Nhe and Hbl

There are similarities between the proteins of Nhe and Hbl (Lund & Granum, 1997). Each operon comprises three genes. Similar sequences are *nheA* and *hblC*, and their structure indicates that they are not coding for any transmembrane helices. *nheB* and *hblD* have two predicted transmembrane helices. *nheC* and *hblA* have one predicted transmembrane helix, in the same position on each gene (Granum *et al.,* 1999). Recent work using two dimensional SDS-PAGE found that only two of components were visible for both Nhe and Hbl complexes which may be because the third components (NheC and HblD – L1) are present in very small quantities (Gohar *et al.,* 2002). There is a fourth gene in the Hbl operon, *hblB*, which is required for the gene to be active, but the mechanism of action is not known.

1.3.2.vi Phospholipase C and sphingomyelinase

Phospholipase C and sphingomyelinase are enzymes involved in synthesis and breakdown of membrane lipids and are associated with virulence. The expression of the genes encoding phospholipase C and sphingomyelinase in *B. cereus* is regulated by PlcR (Lereclus *et al.,* 1996; Granum *et al.,* 1999; Gohar *et al.,* 2002). However, in *B. anthracis* the genes are not expressed due to inactivation of a reduced *plcR* gene (Pomerantsev *et al.,* 2003). The ability of *B. thuringiensis* strains to develop in insects is thought to be associated with production of phospholipases (Lereclus *et al.,* 1996).

Phospholipase C splits the phosphodiester linkage between the hydrophilic polar head group of the membrane lipid and the hydrophobic hydrocarbon (fatty acid) tail (Stryer, 1995 a). *B. cereus* produces three types of phospholipase C, each with a different mechanism (Granum, 1994; Kotiranta *et al.,* 2000). Phosphatidylinositol hydrolase (a non-metalloenzyme) splits phosphatidylinositol and its glycosylated forms. Phosphatidylcholine hydrolase (a stable metalloenzyme, requiring zinc and calcium ions) hydrolyses phosphatidylcholine, phosphatidylethanolamine and phosphatidylserine. Sphingomyelinase (a stable, haemolytic metalloenzyme requiring magnesium ions) is active against sphingomyelin. Cereolysin, a thiol-activated protein is encoded in tandem genes for phosphatidylcholine phospholipase (*cerA*) and sphingomyelinase (*cerB*) (Gilmore *et al.,* 1989; Granum, 1994). The two genes are not transcribed together, but expression of both genes is needed for effective cytolysis and the product is referred to as *cereolysin AB*.

1.3.2.vii CytK

An outbreak of food poisoning in France in 1998 resulted in three deaths (Lund *et al.*, 2000). A symptom of the disease was diarrhoea, and a strain of *B. cereus* was isolated from patients. The *B. cereus* strain did not produce Hbl or Nhe, but was haemolytic when tested with bovine and rabbit cells. A toxin, CytK was isolated by SDS-PAGE and deduced to be a ß-barrel channel-forming toxin, similar to a necrotic enterotoxin from *C. perfringens*. CytK also inhibited protein synthesis in Vero cells. CytK consists of a single necrotic and haemolytic protein of 34 kDa (Lund *et al.*, 2000).

1.3.2.viii PlcR

Many of the genes associated with virulence in the *B. cereus* group have been found to be controlled by a regulator, PlcR, which is similar to regulator proteins found in *Lactobacillus* and *Bacillus stearothermophilus*, PreL and NprA respectively (Lereclus *et al.*, 1996). The recognition site for PlcR is TATGNANNNNTNCATA, and is found 30 to 200 bases upstream of the start codon for the controlled genes. All of the genes regulated by PlcR seem to code for extracellular proteins and include phospholipase C, the haemolytic enterotoxin Hbl and the non-haemolytic enterotoxin Nhe (Agaisse *et al.*, 1999). From this work, it was concluded that *B. thuringiensis* strains were able to produce the same PlcR- regulated toxins as *B. cereus.*

Proteomic studies have been carried out to determine the scope of PlcR regulation. Two-dimensional electrophoresis of proteins from *B. cereus* ATCC 14579 with the *plcR-gene* intact and with the *plcR-gene* disrupted found that the extracellular virulence factors completely or partially controlled by PlcR account for more than 80% of the extracellular proteins in *B. cereus* ATCC 14579. More than half (54%) of the proteins produced by *B. cereus* ATCC 14579 were not present in the strain with the *plcR* gene disrupted. These proteins were proteases (mainly metalloproteases) enterotoxins and phospholipases. From the recently sequenced genome of the type strain of *B. cereus* ATCC 14579, it was estimated that PlcR may be regulating over 100 genes (Ivanova *et al.*, 2003). Interestingly, none of the PlcR-controlled virulence proteins were present in the extracellular proteome of *B. subtilis* (Gohar *et al.*, 2002).

One of the characteristic properties, shown in Table 1.1, which distinguishes *B. anthracis* from *B. cereus* is that *B. anthracis* is non-haemolytic on blood agar and *B. cereus* is haemolytic. The non-haemolytic activity of *B. anthracis* could be caused by a mutation or disruption in the *plcR* gene. The *plcR* gene was found to be present in *B. anthracis,* but a search in the GenBank protein databases did not find any homologous amino acid sequences, suggesting that the *plcR* gene may not be functional. Alignment of the *B. anthracis, B. cereus* and *B. thuringiensis* PlcR amino acid sequences subsequently showed that the sequence in *B. anthracis* is reduced and transcriptionally inactive (Agaisse *et al.,* 1999).

1.3.2.ix Presence of enterotoxin genes in the *B. cereus* group

B. cereus is a common cause of food poisoning and as many cases are undiagnosed, the true extent of the infection is unknown (Kotiranta *et al.,* 2000). The formation of spores allows the bacteria to survive heating and pasteurization. An additional problem is that some strains can survive and grow at temperatures of 4 to 7 ºC and may cause problems with refrigerated foods. Diarrhoeal enterotoxins may be produced by *B. thuringiensis* strains isolated from commercial insecticides. *B. thuringiensis* is phenotypically similar to *B. cereus*, and food poisoning incidents associated with *B. cereus* may actually be due to *B. thuringiensis* through mis-identification. *B. thuringiensis* spores and d-endotoxin crystals have been tested for toxicity towards mammals, but no tests were carried out on the germinated spores and the vegetative cells that they have produced. *B. thuringiensis* strains isolated from ten commercial insecticides reacted positively for diarrhoeal enterotoxin when assayed with Tecra VIA immunoassay kit for NheC, although the amount of toxin produced was lower than that produced by a *B. cereus* control (Damgaard *et al.,* 1995).

A study of *B. cereus* strains isolated from industrial and environmental sources was carried out to investigate food poisoning potential (Pirttijärvi *et al.,* 1999). Ribotyping showed that the emetic toxin producing strains were a clonal group and that some ribotype patterns were similar for strains isolated from milk and dairy processes. Strains obtained from machinery, paper and packaging board did not usually share ribotype patterns, indicating a diverse population with different origins. Strains from industrial and environmental origins were tested for Hbl by BCET-RPLA (reversed passive latex agglutination) and ELISA-VIA (visual immunoassay) to detect Nhe. Thirty strains out

of 56 were found to produce Hbl, 24 out of 44 produced Nhe and 16 out of 43 produced both enterotoxins.

Further studies have found that the enterotoxin genes are common among *B. thuringiensis* serovars and that many strains are able to inhibit protein synthesis in Vero cells (Gaviria *et al.,* 2000). Another study compared 23 strains from the *B. cereus* group (16 *B. thuringiensis*, 4 *B.cereus* and 3 *B. anthracis*) for a wide range of virulence factors. The study found that while *B. anthracis* could be distinguished from *B. cereus* and *B. thuringiensis*, there were no clear distinctions between *B. cereus* and *B. thuringiensis* strains (Guttmann *et al.,* 2000).

Fifty *B. weihenstephanensis* strains were tested for the presence of toxin genes using toxin kits. Most of the *B. weihenstephanensis* strains (72%) were not cytotxic, but 23 strains were tested by PCR and half were found to have the *hbl* gene. All were found to have the *nhe* gene, but very few were positive for *cytK* (Stenfors *et al.,* 2002).

1.4 Classification of *B. thuringiensis*

The subspecific classification of *B. thuringiensis* strains is based on flageller (H) antigen serotyping (de Barjac & Bonnefoi, 1962; Lecadet *et al.,* 1999). Each serotype is given a name and serotype number and while this system has been very useful, unfortunately serotype does not always relate to toxicity.

In 1989, Höfte and Whiteley proposed a classification system for crystal proteins based on their structure and host range, using 14 crystal protein types. This scheme was revised in 1998 to accommodate over 100 crystal protein types (Schnepf *et al.,* 1998). An attempt to correlate 'H' antigen serotyping with crystal protein nomenclature is given in Table 1.2.

1.4.1 'H' antigen serotyping

The specificity of antigen-antibody reactions has been used to classify bacterial strains into groups or varieties within a species. Serotypes are differentiated by agglutination tests with specific antisera. Three types of surface antigens have been used in serotyping

schemes, the O antigens are the polysaccharides in the outer layer of the cell wall, the K antigens are capsular polysaccharides and the H antigens are flagellar proteins. The Kaufman-White classification of Salmonellae was developed in the 1930s using O and H antigens. Hundreds of serotypes have been identified, and the system has been very useful in epidemiological and taxonomic studies (Stanier *et al.*, 1977; Fitzgerald *et al.*, 2003). Similar O and H serotyping schemes have been developed for *Escherichia coli* and K antigen typing has been useful in identifying diarrhoeal strains of *E. coli* in calves (K99 antigen) (Myers & Guinee, 1976).

The H-serotyping method of classifying *B. thuringiensis* strains is simple and effective. However, there are some problems: - some strain are non-motile and lack flagellae, other strains autoagglutinate and are not typeable, a further difficulty is that motile *B. cereus* strains have antigens that react with *B. thuringiensis*-specific sera. Approximately 70 % of *B. cereus* strains isolated from the environment (leaves, soil and faeces) were seropositive when tested with *B. thuringiensis* reference antisera (Shisa *et al.*, 2002).

The International Entomopathogenic Bacillus Centre (IEBC), at the 'Institut Pasteur' maintains a collection of 3,493 strains of *B. thuringiensis* grouped into 69 serotypes on the basis of their flagellar H antigens (January 1999). This technique was developed by de Barjac and Bonnefoi in 1962, and updated by Lecadet *et al*., (1999). Some of the *B. thuringiensis* serotypes of common strains are included in Table 1.2.

1.4.2 *B. thuringiensis* crystal proteins.

The protein crystals are produced by *B. thuringiensis* strains within the mother cell in association with sporulation. They are usually released separately as the bacterial cell lyses but in some cases the crystal remains attached to the spore enveloped within the exosporium e.g. *B. thuringiensis* var. finitimus (Debro *et al.*, 1986). The crystals, which are protoxins of 27 to 140 kDa, are consumed by insect larvae, and dissolve in the alkaline pH conditions found in the mid-gut of insects. Insect proteins hydrolyse the protoxins, into smaller toxic polypeptides. A diagram of the structure of the protoxin Cry3A is shown in Figure 1.4.

The toxin creates pores in the insect's gut cell membrane, disturbing the osmotic balance (Hofte *et al.*, 1989; Aronson & Shai, 2001). Research into *B. thuringiensis* toxin activity has focused on Lepidopteran (Cry1) and Coleopteran (Cry3) toxins. The structures of Cry1, Cry1Aa and Cry3A are similar although there is only 25 to 30 % amino acid homology.

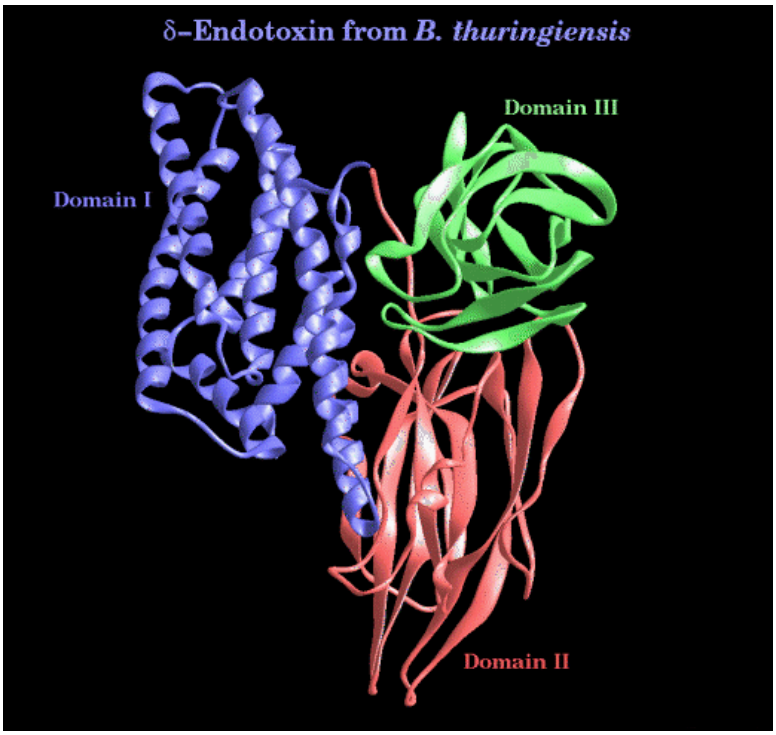


Figure 1.4 Diagram of the structure of *B. thuringiensis* delta-endotoxin Cry3A. Available from 'The Microbial World' at http://helios.bto.ed.ac.uk/bto/microbes/bt.htm

The delta-endotoxin of *B. thuringiensis* has three domains. Domain I consists of seven a-helices which create the pore in the gut cell membrane of the insect larvae. The central helix, a-7 is responsible for opening up Domain I after the toxin has bound to the receptor. Domain II is made up of three ß-sheets that may bind to receptors in the gut membrane. Domain III is a ß-sandwich, which is thought to protect the C-terminus of the toxin against the action of proteases. Domain III also controls ion channel activity (Schnepf *et al.*, 1998).

B. thuringiensis crystal proteins have been used commercially as insecticides for over thirty years, and have been cloned into plants to protect them against insect attack (Schnepf *et al.*, 1998). The crystals are usually specific to particular insect species, and are considered to be harmless to other insects, birds and mammals.

1.4.2.i Classification of *B. thuringiensis* using crystal proteins

In 1989, Hofte and Whiteley developed a classification system for *B. thuringiensis* strains based on crystal protein structure deduced from DNA sequences and host range. For some *B. thuringiensis* strains, the host susceptible to the crystal toxin is unknown. A summary of the Hofte and Whiteley classification is shown in Table 1.2, with serotypes included. Nucleotide sequences for 42 *B. thuringiensis* crystal protein coding genes were allocated to 14 groups: – 13 of these were further divided into four classes, with associated sub-classes. The four classes were CryI – *Lepidoptera* specific, CryII – *Lepidoptera* and *Diptera* specific, CryIII – *Coleoptera* specific and CryIV – *Diptera* specific. A further class, CytA, was proposed for the gene encoding a 27-kDa protein from *Bt* israelensis which was found to have cytolytic activity against vertebrate and invertebrate cells.

CryI proteins (130 to 140 kDa) are bipyramidal in shape, and specific to insects of the order *Lepidoptera*; c*ryI* genes can be separated from the other *cry* genes on the basis of sequence homology (less than 50% amino acid similarity). CryIA(a), cryIA(b) and cryIA(c) have more than 80% amino acid similarity, and form sub-groups. Polyacrylamide gel electrophoresis and immunological tests using polyclonal antisera against crystal proteins showed that some *B. thuringiensis* strains produced two or three different proteins in their crystals e.g *B. thuringiensis* kurstaki HD-1 contains CryIA(a, b and c) (Hofte *et al.*, 1989).

CryII genes encode 65-kDa proteins, which are cuboidal. CryIIA has been found to be active against *Manduca sexta* (Lepidoptera) and *Aedes aegypti* (Diptera). CryIIB is only toxic to *Lepidoptera*, but is classed as CryII because the structures are similar, with only limited homology to the other crystal proteins. CryII genes have been found in several serotypes of *B. thuringiensis* e.g. kurstaki, thuringiensis, tolworthi and kenyae (Hofte *et al.*, 1989).

CryIII proteins are specific to Coleoptera insect larvae, and have been isolated from *Bt* tenebrionis and san diego. The rhomboidal crystals contain a single 72-kDa protein, which was found to be identical from all strains.

CryIV protein genes are Dipteran specific. Four CryIV proteins (A 135-kDa, B 128-kDa, C 78-kDa, and D 72-kDa) and CytA were recovered from a 72-mDa plasmid found in *Bt* israelensis. The crystals assemble together in an ovoid structure. These toxins are of particular importance as they are toxic to some species of mosquito larvae. All four crystals are required for full toxicity to be expressed (Hofte & Whiteley, 1989).

The *cytA* gene encodes a 27-kDa protein, but has no similarity to the *cry* genes. CytA protein has cytolytic properties and is unusual in that it is active against cells from invertebrate and vertebrate animals, including mammalian erythrocytes (Thomas & Ellar, 1983; Schnepf *et al.,* 1998). The toxin structure of CytA is a single domain, with two outer a-helix layers wrapped around a mixed ß-sheet. It has been suggested that the toxic effect of Cyt proteins is similar to the disrupting action of detergents on membranes (Schnepf *et al.,* 1998).

Other insecticidal proteins are secreted during vegetative growth by some strains of *B. thuringiensis* and are called VIPs (vegetative insecticidal proteins). The *vip1A* gene encodes an 80-kda protein and has sequence similarity to the protective antigen of *B. anthracis* toxin (Schnepf *et al.,* 1998)

.

Table 1.2 Crystal Proteins of *B. thuringiensis* – adapted from Hofte & Whiteley (1989), Crickmore *et al.*, (1998), Priest & Goodfellow (2000), BGSC catalog (1999).

Crystal protein toxin				
OLD name Hofte and Whiteley 1989	**NEW name** Crickmore et al 1998	Host Range (Class – Insecta) ORDER	*B. thuringiensis* serovar	H antigen serotype
CryIA (a)	Cry1Aa	Lepidoptera	kurstaki sotto galleriae entomocidus aizawai morrisoni	3a3b3c 4a4b 5a5b 6 7 8a8b
CryIA (b)	Cry1Ab	Lepidoptera	thuringiensis kurstaki aizawai	1 3a3b3c 7
CryIA (c)	Cry1Ac	Lepidoptera	kurstaki kenyae	3a3b3c 4a4c
CryIB	Cry1Ba	Lepidoptera	thuringiensis	1
CryIC	Cry1Ca	Lepidoptera	kenyae entomocidus aizawai	4a4c 6 7
CryID	Cry1Da	Lepidoptera	aizawai	7
CryIE	Cry1Ea	Lepidoptera	kenyae tolworthi	4a4c 9
CryIIA	Cry2A	Lepidoptera / Diptera	kurstaki	3a3b3c
CryIIB	Cry2Ab	Lepidoptera	kurstaki	3a3b3c
CryIIIA	Cry3Aa	Coleoptera	tenebrionis sandiego	8a8b 8a8b
CryIIIB	Cry3Ba	Coleoptera	tolworthi	9
CryIVA	Cry4Aa	Diptera	israelensis	14
CryIVB	Cry4Ba	Diptera	israelensis	14
CryVA (a)	Cry5Aa	Lepidoptera / Coleoptera	darmstadiensis	10a10b
CryIH	Cry9Ca	Lepidoptera	tolworthi	9
CryIVC	Cry10Aa	Diptera	israelensis	14
CytA		Diptera / cytolytic	israelensis morrisoni	14 8a8b

Crystal protein coding genes are located on plasmids, which may be transferable between strains of *B. thuringiensis*. The accumulation of genes coding for several toxins would obviously be of considerable benefit to the bacterium as the host range would be expanded.

A more recent system for classifying crystal proteins of *B. thuringiensis* has been developed (Crickmore *et al.*, 1998). The proteins are classified according to evolutionary divergence, as measured by sequence homology. The sequences have been used to construct the phylogenetic tree shown in Figure 1.5. This system of nomenclature has been designed to accommodate the increasing number of Cry proteins and to address some anomalies in the earlier system. Table 1.2 shows how the new nomenclature relates to the Hofte & Whiteley scheme. Roman numerals have been replaced by Arabic numerals in the primary rank, to allow for the addition of new proteins. Vertical lines drawn through the tree in Figure 1.5 define the ranks. Each toxin has a name that relates to its position on the tree. A new toxin will be given a new primary rank if it joins the tree to the left of the 45% amino acid sequence identity line. If the toxin enters the tree between the left and central boundaries, a new secondary rank (uppercase letter) will be assigned. The tree boundaries shown in Figure 1.5 were designed to retain most of the older names, to avoid confusion.

The Bacillus Genetic Stock Centre (BGSC) maintains a standing committee on Crystal Protein Nomenclature. New gene or protein sequences must be deposited in a public database, such as GenBank and also to the BGSC. An appropriate name will be suggested by the director of the BGSC. A comprehensive list and other information is available from http://www.biols.susx.ac.uk/Home/Neil_Crickmore/Bt/.

The underlying idea of the revised system of Crickmore *et al.*, (1998) *w*as to include information about the evolution of the crystal protein coding genes in the nomenclature.

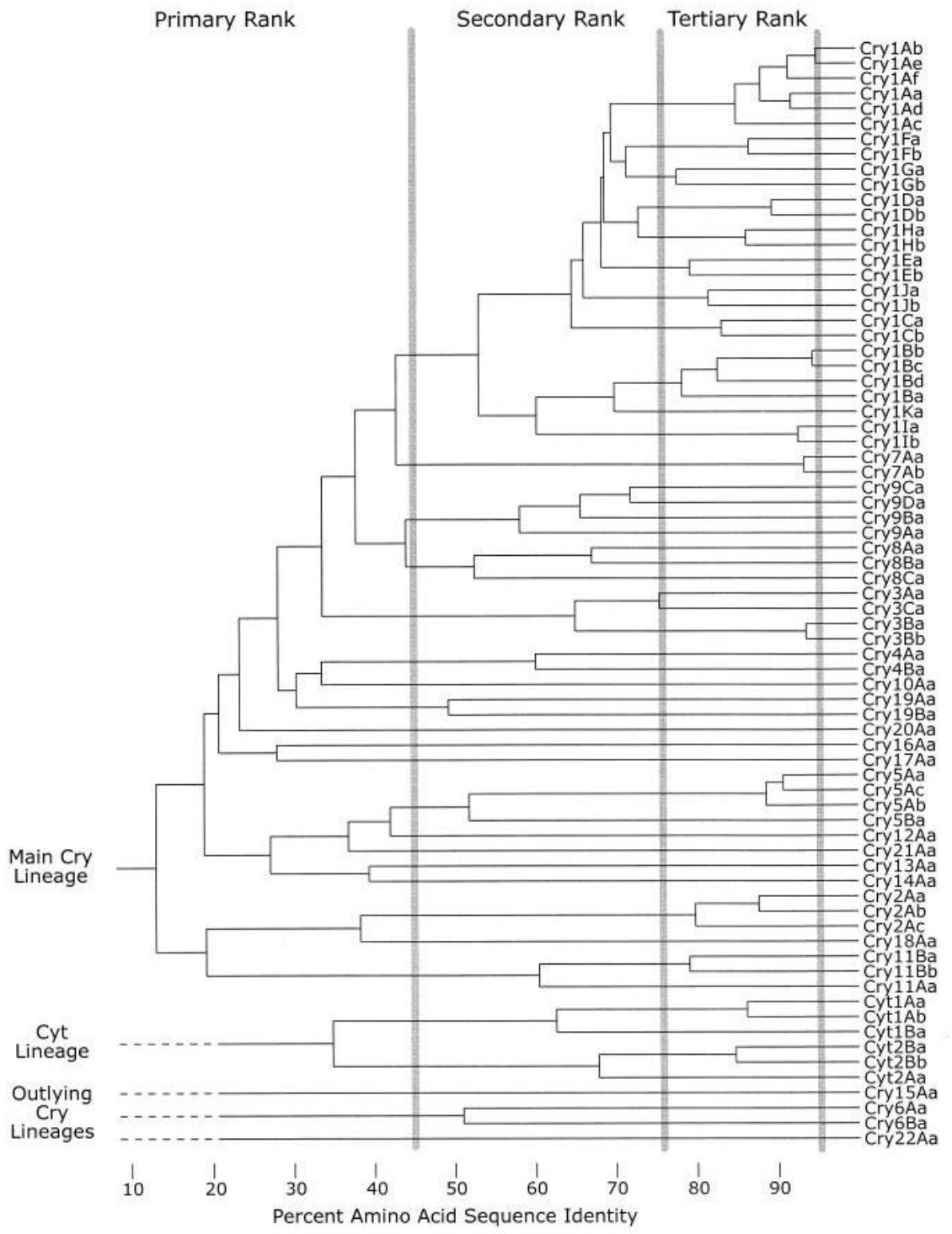


Figure 1.5 Phylogram demonstrating amino acid sequence identity among *Cry* and *Cyt* proteins. This is a Neighbour Joining tree produced from a ClustalW alignment and visualized using TREEVIEW. The grey vertical bars show the four levels of nomenclature ranks. The lower four lineages showed no conserved sequence blocks in alignments, and the nodes have been replaced with dashed horizontal lines (Crickmore *et al.*,1998).

1.5 Analytical methods used for species and subspecies classification of the *B. cereus* group

'Bacterial classifications are devised for microbiologists, not for the entities being classified. Bacteria show little interest in the matter of their classification'
(Brenner *et al.*, 2001).

The *B. cereus* group is at present divided into six species. The three major species are *B. anthracis, B. cereus and B. thuringiensis*, while the three minor species are *B. mycoides, B. pseudomycoides* and *B. weihenstephanensis*. All species are very closely related and there have been many proposals to unite them into a single species (Gordon *et al.*, 1973; Kaneko *et al.*, 1978; Helgason *et al.*,1998). Conversely, many analytical methods have been used to try to separate the *B. cereus* group into the six constituent species and some of these studies are outlined below.

1.5.1 Phenotype

Many studies have been undertaken to try to sub-divide the *B. cereus* group using phenotypic criteria. The criteria examined have included colony morphology, cell morphology, including spore shape and position. Physiological criteria have also been investigated, such as the degradation of various materials including chitin, lecithin, casein and gelatin. Other tests investigated have covered antibiotic resistance, acid production from sugars and single carbon sources for energy and growth such as acetate, lactate, citrate or malonate. The ability to grow in different sodium chloride concentrations, at varying temperatures, or anaerobically, have also been studied. A numerical taxonomic study, involving 118 tests was carried out on 368 *Bacillus* strains including 58 strains from the *B. cereus* group (excluding *B. anthracis*) (Priest *et al.*, 1988). The strains related to *B. cereus*, were recovered in nine sub-clusters, some of which were mixed, consisting of *B. thuringiensis* and *B. cereus* strains. Seven food-poisoning *B. cereus* strains were recovered together and 12 strains of *B. thuringiensis* formed another separate group.

The distinctive rhizoidal colony growth distinguishes *B. mycoides and B. pseudomycoides* from the other species. However, no consistent features, other than crystal protein synthesis, to distinguish between *B. cereus* and *B. thuringiensis* were found.

1.5.2 DNA hybridization

Phenotypic methods of identifying bacteria are sometimes ambiguous and the 'International Committee on Systematics of Prokaryotes (ICSP), formerly the International Committee on Systematic Bacteriology (ICSB)' describe DNA:DNA hybridization as a useful tool to determine relatedness among bacteria (Brenner *et al.,* 2001). DNA hybridization is carried out by denaturing native double-stranded DNA (by heating), followed by cooling the DNA quickly. The DNA is incubated with a sample of denatured DNA or RNA from another bacterium, at an appropriate annealing temperature and will form heteroduplexes with any complementary sequences. Strains with approximately 70% or greater DNA-DNA relatedness and a thermal stability within 5ºC of the thermal stability of the homologous duplex are considered to be the same species A feature of such a method using DNA relatedness is that the effects of unusual reactions, pathogenicity and plasmids are very small. Since limited DNA hybridization studies have found no clear distinctions between *B. anthracis, B. cereus,* and *B. thuringiensi,* revised groupings have been proposed (Somerville *et al.,* 1972; Kaneko *et al.,* 1978; Nakamura, 1994).

1.5.3 rRNA methods – sequencing and ribotyping

1.5.3.i Analysis of DNA sequences coding for rRNA genes

Ribosomal RNA genes are highly conserved and widely recognized as good examples of 'molecular clocks' (Ash *et al.,* 1991; Ash *et al*., 1992). Analysis of genes encoding for 16S and 23S ribosomal RNA is well established as a tool for classifying prokaryotes and identifying species. Ash and her colleagues found the 16S rRNA and 23S rRNA sequences of *B. anthracis* (Sterne strain) and *B. cereus* emetic toxin producing strain NCTC 11143 to be identical. However, the 16S rRNA sequences for *B. mycoides* and *B. thuringiensis* differed from each other and from strains of *B. anthracis /B. cereus* by

four to nine nucleotides. Nevertheless, this is all within the range expected of a single species.

The intergenic spacer region (ISR) between the 16S and 23S rRNA genes is often more variable than the actual genes. By sequencing this region, sufficient information was provided to distinguish *B. anthracis* strains from *B. cereus* strains although the difference was only 1 nucleotide (Harrell *et al.,* 1995). The method could not distinguish between the two *B. anthracis* strains tested. However, there were 13 nucleotide differences between *B. anthracis* and *B. mycoides*. In the same study, PFGE was carried out, using the restriction enzymes *Not*I, *Sfi*I, and *Sma*I to digest chromosomal DNA. The three *B. anthracis* strains tested gave identical band patterns. Strains of *B. cereus* and *B. mycoides* could be distinguished from *B. anthracis* and from each other, supporting the results obtained by comparing sequences from the ISR region.

A study using mung bean nuclease treatment and sequencing of the 16S–23S rRNA gene intergenic transcribed spacer region of strains from the *B. cereus* group has been carried out (Cherif *et al.,* 2003). A phylogenetic tree based on the long ITS sequences containing tRNA genes identified four groups of strains: – (1) *B. cereus / B. thuringiensis*, (2) *B. anthracis*, (3) *B. mycoides / B. weihenstephanensis* and (4) *B. pseudomycoides*.

An extensive DNA sequencing study compared 183 strains from the *B. cereus* group for variations in 16S rRNA gene signature codons and 74 strains for differences in 23s rRNA sequences. (Bavykin *et al.,* 2004). Based on 16S rRNA sequence homology, the *B. cereus* group was found to divide into seven sub-groups: - Anthracis, Cereus A and B, Thuringiensis A and B and Mycoides A and B. All strains of *B. anthracis* were found to be homogeneous. *B. cereus* strains were recovered in six of the seven groups and *B. thuringiensis* in three of the groups. The cluster groups determined using sequence data for 23S rRNA genes agreed generally with the groupings based on 16S rRNA genes (Bavykin *et al.,* 2004).

1.5.3.ii Ribotyping

Ribotyping is a method of typing strains of bacteria by digesting chromosomal DNA with restriction enzymes and transferring the fragmented DNA to a membrane by Southern blotting. The membrane bound DNA is then hybridized to a prokaryotic 16S rRNA gene probe. Ribotype patterns are produced which reflect the distribution of restriction enzyme sites within and between the 16S rRNA operons.

An analysis of rRNA operon restriction fragment length polymorphisms (RFLPs) was used to compare patterns from *B. anthracis, B. cereus, B. thuringiensis* (43 strains from ten serovars) and *B. mycoides* (Priest *et al.*,1994). The ribotype patterns from the four *B. anthracis* strains were found to comprise a distinct group and RFLPs for *B. thuringiensis* strains correlated well with their serotype, suggesting a clonal structure, while the three *B. cereus* strains were dispersed on the resultant phylogenetic tree.

In a more extensive study, rRNA RFLPs of 86 strains of *B. thuringiensis* isolates, including some strains with the same serotype were generated. DNA was digested with *Eco*R1 and *Hind*III and probed with a labelled DNA fragment constructed from *B. subtilis* 16S rRNA (Joung & Côte, 2001 a). The use of two restriction enzymes proved to be more discriminatory in that the *Bt* sandiego and *Bt* tenebrionis had the same band pattern with *Hind*III, but different patterns with *Eco*RI. However, *Hind*III digests produced different patterns for *Bt* kurstaki strains HD-1 and HD-73 while *Eco*RI gave the same pattern for both strains. A phylogenetic tree based on the 16S rRNA gene fingerprints showed that the strains clustered in four groups. Strains of *Bt* kurstaki, tolworthi, canadensis and aizawai were found in group I; *Bt* israelensis, morrisoni, sandiego, tenebrionis and darmstadiensis were in group IV. (Groups II and III contained *Bt* monterrey, andaluciensis, seoulensis and cameroon which were not examined in this thesis). Further work by the same authors compared the 16S rRNA ribotyping results with 23S and 5S rRNA RFLP ribotyping. Since the 23S and 5S genes contain about twice as many bases as the 16S gene, they should provide better discrimination. The data from 16S, 23S and 5S rRNA ribotyping were combined and a phylogenetic tree constructed which was considered to be more robust as more informaton was included. Eight groups were recovered from the combined data. Group I contained *Bt* kurstaki, aizawai, pakistani and others, whereas *Bt* israelensis, morrisoni, sandiego and tenebrionis were in a different group, in this case group II. DNA homology between

strains of *B. thuringiensis* was high (>91%) and it was concluded that a method of sub-classification based on RFLP ribotyping could be very useful and probably more informative than 'H' antigen serotyping. However, more strains from within each serotype need to be examined (Joung & Côte, 2001 b).

1.5.4 Random Amplified Polymorphic DNA (RAPDs)

The RAPD technique involves using a single short primer in a PCR reaction to amplify sections of DNA. Following electrophoresis, band patterns are revealed and strains producing identical bands are likely to have identical or very similar genomes.

B. thuringiensis serovars could be distinguished from each other and from *B. cereus* using this technique. Closely related commercial strains of *B. thuringiensis* from biopesticides were distinguishable by differences in band patterns (Brousseau *et al.*, 1992). Brousseau and colleagues mention that the amount of DNA and the type of Taq polymerase used affected the band patterns. A study of 126 strains of *B. thuringiensis* by RAPD-PCR found evidence that RAPD band patterns correlated with some serovars (e.g. kurstaki, sotto, thuringiensis), suggesting that clonal groups have arisen in *B. thuringiensis* (Gaviria & Priest, 2002).

Another study of 101 strains from the genus *Bacillus*, including 61 strains from the *B. cereus* group was carried out using RAPDs (Daffonchio *et al.*, 1999). A characteristic band of 850-bp was amplified from all strains of the *B. cereus* group but not in strains from other *Bacillus* species. Southern hybridisaton with a digoxigenin (DIG) labelled probe made with DNA excised from the 850-bp RAPD band hybridised only with *B. cereus* group strains. The 850-bp fragment was amplified by PCR and products were obtained from *B. cereus* group strains, including *B. anthracis* strains that had been cured of one or both pXO plasmids, showing that the fragment is probably situated on the chromosome, rather than on a plasmid. The DNA was digested with the restriction enzyme *Alu*I. A unique restriction profile was found in the *B. anthracis* strains separating *B. anthracis* from the other *B. cereus* group species.

1.5.5 Typing methods using restriction enzyme digests

1.5.5.i Pulsed field gel electrophoresis (PFGE)

The technique of PFGE can separate large DNA fragments of up to 800 kb. Chromosomal DNA is digested with restriction enzymes which will only recognize a few cutting sites in the chromosome. After electrophoresis, the resultant band patterns are compared. PFGE was carried out to try to assess the genomic variation between *B. cereus* and *B. thuringiensis* strains (Carlson *et al.*, 1994). No consistent differences were found between strains recognized as *B. cereus* or *B. thuringiensis*, and it was concluded that they belonged to the same species. A PFGE study of 70 strains of *B. thuringiensis* from 21 serovars found that some serovars were clonal in structure e.g. *Bt* darmstadiensis, kurstaki, thuringiensis and sotto, while others were heterogeneous e.g. *Bt* canadensis (Gaviria *et al.*, 2003).

1.5.5.ii PFGE plus chromosome mapping

The technique of chromosome mapping was applied to *B. cereus* ATCC 14579 and *B. thuringiensis* canadensis (Carlson, 1996). The chromosomes were digested with three restriction enzymes, *Asc*I, *Not*I and *Sfi*I, and the fragments separated by PFGE, to produce fingerprints of the two strains, which proved to be very similar. Cross-hybridization of 11 *Not*I fragments from *B. cereus* ATCC 14579 and 10 from *Bt* canadensis showed that the fragments from these two chromosomes shared DNA homology. Additionally, gene probes were used to build up physical and genetic maps of the two strains and as the physical maps were almost identical, they were plotted as one map. Interestingly, *Bt* canadensis and *B. cereus* ATCC 14579 were found to have similar *hblA* sequences although an enterotoxin assay result was negative for the canadensis strain (*hblA* codes for the binding component of the haemolytic enterotoxin Hbl).

1.5.5.iii Fluorescent amplified fragment length polymorphisms (AFLPs)

To generate an AFLP, genomic DNA is cut with restriction enzymes such as *Eco*RI and *Mse*I to produce fragments of DNA. Double stranded adaptor molecules are ligated to the ends of the DNA fragments to create sites that PCR primers can recognize and anneal to (Vos *et al.*, 1995). Following PCR, the products are separated by electrophoresis to reveal a pattern (fingerprint) for each strain with closely related

strains having similar banding patterns. Typing of 21 strains of *B. cereus* by amplified fragment length polymorphisms found 16 patterns (Ripabelli *et al.,* 2000).

In fluorescent AFLP, the *EcoRI* primer is labeled with a fluorescent dye (FAM). The resulting AFLP products are mixed with DNA size standards (labeled with another fluorescent dye (TAMARA)) and loaded onto a sequencing gel. 'Genescan' software (Applied Biosystems) computes the length of the sample fragments by comparing them with the DNA size standards. The relative position of the fragment on a gel is translated into a peak of a particular height at a specific location. A comparison of 154 Norwegian environmental isolates of *B. cereus / B. thuringiensis* by fluorescent AFLP, recovered five major clusters. *B. anthracis* strains (2) were closer to *B. cereus* human pathogens from periodontitis than they were to the strains isolated from soil (Ticknor *et al.,* 2001). An extensive fluorescent AFLP study examined over 300 diverse strains from the *B. cereus* group including *B. anthracis* (Hill *et al.,* 2004). Three major clusters with ten branches were resolved: - cluster 1 contained branches A, B and C., cluster 2 was made up of branches D, E, F and G while cluster 3 contained branches H, J and K. All of the *B. anthracis* strains were together on branch F confirming their status as a clonal group with peculiar pathogenic properties. *B. cereus* and *B. thuringiensis* isolates were dispersed over all branches but 22 out of 42 pathogenic *B. cereus* strains were recovered in cluster 2. The intermixing of *B. cereus* and *B. thuringiensis* strains over all three clusters led the authors to suggest that the insecticidal or pathogenic properties arose after the division into 3 clusters probably by plasmid transfer. Branch A in cluster 1 was composed entirely of *B. thuringiensis* strains and branch C in the same cluster was predominantly *B. thuringiensis,* indicating the probable emergence of clonal groups of insect pathogens.

In summary, most of the research carried out has reached similar conclusion: -

- *B. anthracis* strains were homogeneous, but could usually be distinguished from other *Bacillus* strains.
- *B. cereus* and *B. thuringiensis* strains were mixed up together in phylogenetic trees.
- *B. mycoides* strains were quite well separated from the other species but actually formed two distinct groups - *B. mycoides sensu stricto* and *B. pseudomycoides.*

1.5.6 Typing methods using DNA sequences of protein coding genes

Molecular typing methods depend on comparing DNA fragments from a single locus or multilocus. When a single gene is used, a rapidly evolving gene is selected (such as a gene coding for cell surface structure), and strains selected for analysis will usually be different from each other (Spratt, 1999). This approach is useful for studying epidemic outbreaks in a hospital or other community, but to investigate the long-term evolution of bacterial populations, a multilocus method is required. Genes that control essential cell metabolism are very stable, and diversify slowly. These 'housekeeping' genes are relatively uniform in sequence and analysis of a single locus would be complicated by recombinational events. Consequently, multilocus approaches are more robust.

1.5.6.i Single gene phylogenetic typing – *gyrB*

DNA sequences from a housekeeping gene have been compared in attempts to differentiate species of the *B. cereus* group. *gyrB,* (topoisomerase type II) was selected as a possible marker gene, and PCR primer sets were designed to amplify fragments of the gene which were specific for *B. cereus, B. thuringiensis* or *B. anthracis*. When 104 strains from the *B. cereus* group were tested with the *gyrB* primers, most amplified the expected product. However, some of the strains classified as *B. cereus* reacted with PCR primers that were designed to be specific for *B. thuringiensis* or *B. anthracis*. Similarly, some *B. thuringiensis* strains (cry^+) and a strain of *B. mycoides* produced the *B. cereus* type of PCR product. The technique of PCR amplification of *gyrB* was applied to test boiled rice for *B. cereus* contamination. Boiled rice was artificially contaminated with *B. cereus,* homogenized and incubated for 15 h in Nutrient Broth. PCR amplification was performed directly on the broth, without DNA extraction indicating the value of this approach (Yamada *et al.,* 1999).

Sequence analysis of *gyrB* has also been used to compare 18 strains of *B. cereus,* five *B. thuringiensis* and reference strains of *B. anthracis, B. mycoides*, *B. pseudomycoides* and *B. weihenstephanensis.* Additionally, seven environmental isolates from spacecraft associated surfaces were included. The gene for 16S rRNA was also sequenced from all strains but the resultant phylogenetic tree showed no cluster grouping at all. The phylogenetic tree generated from a 1.2-kb sequence of the *gyrB* gene, however, produced four distinct groups. Group 1 included *B. anthracis* and ten *B. cereus* strains.

Group 2 was made up of seven *B. cereus* and one *B. thuringiensis* strain. Group 3 was solely *B. thuringiensis* and group 4 was a single *B. mycoides*. DNA:DNA reassociation values from hybridization analyses recovered the same four groups with the exception of the *B. thuringiensis* group. *gyrB* sequence analysis and DNA hybridization both discriminated strains below species level (La Duc *et al.,* 2004).

1.5.6.ii Multilocus Sequence Typing (MLST)

Multilocus sequence typing was developed in 1998. From the methods available before that date, multilocus enzyme electrophoresis (MLEE) had proved to be the most useful for identifying lineages within bacterial populations and for epidemiological studies. However MLEE studies carried out on strains from the *B. cereus* group have been inconclusive. A study carried out MLEE on 36 strains of *B. anthracis*, *B. cereus* and *B. thuringiensis*, reported that the species appear to be genetically indistinguishable. The authors suggested that horizontal transfer of plasmids could possibly provide virulence characteristics for the bacteria, but it may be that additional genetic features are required (Helgason *et al.*, 2000 a). Another MLEE study found that in sympatric populations, gene flow was higher between strains classified as belonging to the species *B. thuringiensis,* or between strains designated as *B. cereus,* than it was between *B. thuringiensis* and *B. cereus* strains. However, recombination between *B. cereus* and *B. thuringiensis* strains was still possible (Vilas-Boas *et al.*, 2002).

The main problem with MLEE and other methods that depend on electrophoretic mobility is that the results are difficult to compare between laboratories. The MLEE method of taking several 'housekeeping genes' from each strain was adopted. PCR was used to amplify 450-bp to 500-bp regions from each gene and the gene fragments were sequenced and compared (Maiden *et al.,* 1998).

The steps required for MLST are as follows: -

1 Select about seven 'housekeeping' genes
2 Sequence the gene fragments on both strands
3 Compare sequences of each fragment with known alleles
4 Give each allele for each locus a number e.g. *glp*-1, *glp*-2 etc
5 The seven alleles define the sequence type (ST)

6 Each ST now has a unique code, e.g. ST 26 (3,2,31,5,16,3,4)

7 Add the data to the MLST database

Housekeeping genes are generally chosen for MLST as they code for essential proteins and would be present in all strains, without insertions or deletions (which would make alignments difficult). When the data obtained from MLEE were compared with MLST data for *Neisseria meningitidis* it was found that resolution using the sequence data from six alleles was equivalent to that obtained using twelve enzymes in MLEE (Maiden *et al.,* 1998). Any nucleotide change in an MLST scheme produces a new ST, but in MLEE about 26 nucleotide changes are needed to produce a new electrophoretic type. (Boyd *et al.,* 1994). Using several genes was necessary because recombination or mutation alters the sequence and could affect the relative position of the strain in a phylogenetic tree if a single gene was used. However, using more than seven genes did not improve the resolution very much, and seven genes have become the standard for most MLST schemes. It was important that the genes were chosen from around the chromosome, to avoid problems with linkage. Nucleotide sequence data is easy to transfer between laboratories all over the world via the internet and the first MLST database was set up for meningococci. The data from MLST were found to be suitable for the study of population structures and to examine the extent of recombination by comparing phylogenetic trees for each allele. Any population with enough nucleotide changes to generate different alleles and sequence types could be typed by MLST (Maiden *et al.,* 1998).

Seven loci are now routinely used for MLST projects, and the technique has been developed for many bacterial species, e.g. *Campylobacter jejuni, N. meningitiditis, Staphylococcus aureus, Streptococcus pneumoniae* and *Candida albicans.* More systems are currently under development and the method is becoming widely established. As the databases expand, more sophisticated modeling of the dynamics of bacterial populations and their evolution will become possible (Urwin & Maiden, 2003).

An MLST scheme for *Streptococcus pneumoniae* has been developed. Some of the strains used had been analyzed before with PFGE and MLEE, and the MLST groupings were consistent with those obtained by the other methods. An unexpectedly high level of congruence was found between serotype and genetic relatedness as determined by

MLST (Enright & Spratt, 1998). There is usually poor congruence between genetic relatedness derived from serotype and MLEE data, as recombinational events alter the serotypes, and may be positively selected by the immune system of the host. The congruence found in the MLST scheme may be due to the recent emergence of clonal complexes. Furthermore, MLST was able to identify antibiotic resistant clones of *S. pneumoniae*, and assigned all of the penicillin resistant clinical isolates to the known resistant clone group.

A MLST system has been developed for *Campylobacter jejuni*, based on 194 isolates which were assigned to 155 STs and grouped into 62 clonal complexes The same allele was present in isolates from different origins and the most likely mechanism for this to happen would be by recombination rather than mutation (Dingle *et al.*, 2001). It was anticipated that the MLST database for *C. jejuni* would be a useful tool for local and global epidemiology. However, an attempt to type *Escherichia coli* 01576:H7 isolates by MLST found insufficient sequence diversity and better separation of strains was obtained by PFGE (Noller *et al.*, 2003). The researchers found that all the 77 strains studied (59 from Pennsylvania and 18 from Minnesota) had identical sequences at all seven loci. This was thought to have happened because the differences between strains of *E. coli* 0157:H7 were due to insertions and deletions, which were localized in specific areas, and were not included in the housekeeping genes selected for MLST.

As yet, there are few MLST schemes for diploid organisms, but a system has been developed for *Candida albicans*. Six loci were investigated and the primers could amplify both homozygous and heterozygous chromosomal alleles but this makes analysis of MLST data more difficult. However, MLST gave high resolution for strain identification and good discrimination even when only three loci were studied (Bougnoux *et al.*, 2002).

The MLST databases are available at http://pubmlst.org. The largest collection of isolates is represented by *N. meningitidis*, which has about 5000 strains. Other large databases, containing over 1000 entries are *C. jejuni, S. aureus* and *S. pneumoniae.* Recent developments to help manage large sample sizes include 'eBurst' software, which uses the allelic profile obtained for each isolate and divides the data into clonal complexes made up of isolates which share identity at six or seven alleles with at least

one other member of the group. The founder of the complex is identified as the genotype that differs from the highest number of other genotypes in the clonal complex at only one locus out of seven. The results are shown graphically (Spratt *et al.*, 2004; Feil & Enright, 2004).

1.5.7 Comparison of whole genome sequences

The genome sequence of *B. anthracis* Ames strain was recently published (Read *et al.*, 2003). Anxiety over the possible use of *B. anthracis* spores in bioterrorist attacks resulted in a large amount of funding being given to *B. anthracis* sequencing projects. There are now complete genome sequences for 15 *B. cereus* group isolates, including the *B. cereus* type strain ATCC 14579 and strain ATCC 10987 - a strain isolated from a study of cheese spoilage in Canada (Rasko *et al.*, 2004). It is anticipated that a comparison of the genome sequences will reveal nucleotide polymorphisms which can be used for strain discrimination of *B. anthracis* and may also elucidate the evolutionary history of the species (Rasko *et al.*, 2005). Other sequenced strains include *B. cereus* G9241 isolated from a patient with anthrax-like symptoms (Hoffmaster *et al.*, 2004), *B. cereus* Zebra Killer (ZK), isolated from a Zebra carcass in Namibia and *Bt* konkukian, strain 97-27 which was isolated from a wounded French soldier in Sarajevo (Hernandez *et al.*, 1998).

A *B. anthracis* DNA microarray was used to investigate comparative genome hybridization with *B. cereus* strains and a *B. weihenstephanensis.* The chromosomal genes were found to be very similar, with a region of conserved genes found near the origin of replication. The differences between *B. anthracis* and the other species of the *B. cereus* group may be due to differences in gene expression, rather than the presence or absence of genes (Read *et al.*, 2003).

In summary, methods for typing bacteria that depend on separation of gene fragments by electrophoresis have proved to be very useful but the results are difficult to compare between laboratories.

A more robust method would be to compare DNA sequence data. At present, whole genome sequencing is laborious, time-consuming and expensive. The sequencing of approximately 500-bp fragments from 7 housekeeping genes provides a quick method for typing bacterial strains quickly and provides unambiguous data which can be readily compared between laboratories all over the world.

1.6 Population Structure

Bacteria divide by binary fission and do not benefit from the mixing of genes that occurs in eukaryotic sexual reproduction. Bacteria however, are able to conjugate and part of a chromosome can be transferred from a donor to a recipient cell. Transducton by phages and natural transformation, in which DNA fragments (resulting from donor cell death) are taken up by the recipient, are also methods which can result in the unidirectional exchange of genetic material between bacteria. Indeed, sometimes, genetic material from a distantly related organism can be acquired in this way (Spratt *et al.*, 2001). Mutations occur naturally by errors in replication, or by chemical and physical mutagens (Brown, 1999). Some mutations are synonymous, having no effect on the bacterial phenotype. Other mutations will alter the type of amino acid produced and the amino acid sequence of the resulting protein. Such non-synonymous mutations are usually harmful to the organism and are fatal or reduce fitness. However, some mutations may be beneficial to the organism: these mutations will be transferred to the progeny of the mutant and may survive and increase in the population by 'Darwinian' natural selection of the fittest. A population where there is no transfer of genetic material, and no mutation, would result in identical individuals or 'clones'.

Analysis of the results of MLEE studies led to the suggestion that some bacterial populations were clonal, e.g. *Salmonella*, while others were panmictic, with high levels of recombination (Maynard Smith, 1993). An intermediate or 'epidemic' type of population was found in *N. meningitidis* where the background population is panmictic, but stable clonal groups are emerging. Figure 1.6 shows a spectrum of clonality in bacterial populations.

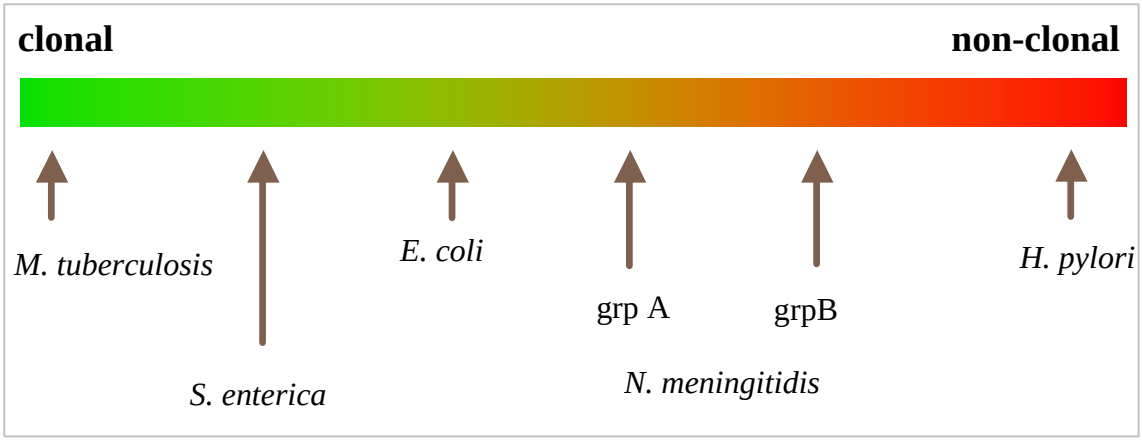


Figure 1.6 Clonal spectrum of bacteria

Maynard Smith calculated that for a population to be totally non-clonal or panmictic, the recombination affecting a gene must occur 2-4 times as often as point mutations, and the index of association will be 0 ($I_A = 0$).

The I_A is a statistical test which attempts to measure the extent of linkage equilibrium within a population by quantifying the amount of recombination among a set of sequences and detecting association between alleles at different loci (Maynard Smith *et al*., 1993). I_A is calculated as follows:

$$\mathbf{I_A = VO/VE - 1}$$

In which, if V0 is the observed variance of K and VE is the expected variance of K, where K is the number of loci at which two individuals differ. If there is linkage equilibrium because of frequent recombination events, the expected value of I_A is zero. Clonal populations are identified by an I_A value that differs significantly from zero.

In non-clonal or panmictic populations, where there are high levels of gene flow, the individual genes are different from each other. When compared by MLEE the enzymes will electrophorese at different rates. When compared by RAPD, the pattern of bands produced will be different and MLST will confirm that the DNA sequences are not the same. There is linkage equilibrium between alleles in a freely recombining panmictic population (Maynard Smith *et al.*, 1993). Clonal groups of bacteria show linkage disequilibrium between alleles because there is no reassortment of mutations through recombination. However, linkage disequilibrium may arise in populations where there is frequent recombination of genes. This can happen if there is a mixture of several populations, each of them recombining within their own population, but not with other populations. Other reasons could be genetic drift or an epidemic population structure.

A high frequency of recombination was found in a study of *Helicobacter pylori*, a bacterium that lives in the gastric mucosa of the human stomach, and is found in half of the population of the world (Falush *et al.*, 2001). Analysis by MLST of 10 genes from 26 pairs of strains (two strains from each patient) found high rates of recombination between unrelated strains that resulted in a panmictic population. At the other extreme, *Mycobacterium tuberculosis* has been found to be highly clonal, with 225 strains isolated in England and Wales forming just four lineages, using MLST of seven genes (Baker *et al.*, 2004). *N. meningitidis* exemplifies an intermediate type, with an ‘epidemic‘ structure (Figures 1.6, 1.7). Calculations have shown that the rate of recombination in *N. meningitidis* is up to twice the rate estimated for *E. coli* (Feil *et al*., 1999).

N. meningitidis is naturally transformable. When the MLST data for 107 strains were analysed, which included several representatives from invasive lineages, a high I_A value of 4.57 was calculated, indicating a clonal population. When the data were re-examined with single representatives from the invasive lineages, an I_A value of 0.65 was observed, indicating a panmictic structure. Such a difference in I_A may be due to the inclusion of many isolates from disease in the sampling, and is characteristic of an epidemic type of population as shown in Figure 1.7 (Holmes *et al.*, 1999).

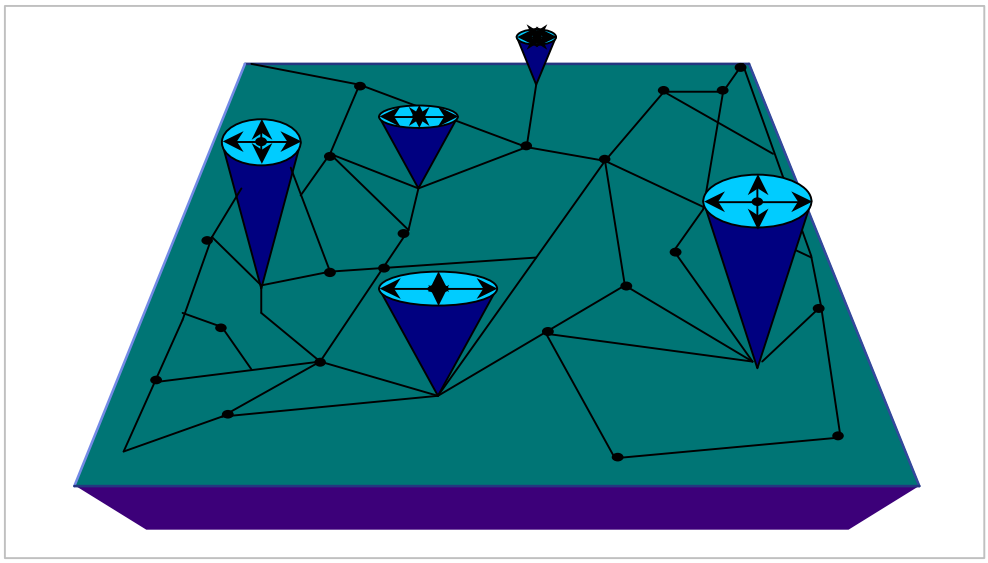


Figure 1.7 'Epidemic' model of clonal expansion in bacteria.
This diagram is from the web page of E. Feil (http://www.bath.ac.uk/bio-sci/feil.htm).

An epidemic population is made up of two distinct populations. There is a panmictic background population, which is freely recombining, and is more appropriately represented by a network than a bifurcating tree. Clonal groups or lineages represented by the cones in Figure 1.7 are emerging from the background. The clones are successful adaptations and reproduce quickly giving rise to epidemic outbreaks.

In summary, much work has been carried out on bacterial population structure, with particular emphasis on the occurrence and distribution of pathogenic types. However, the population structure and evolution of the six species within the *Bacillus cereus* group remains elusive.

1.7 AIMS OF THIS WORK

For many years, controversy has surrounded the species status given to the six taxa comprising the *B. cereus* group. The main aim of this project was to develop an MLST scheme for these bacteria in order to clarify the phylogenetic structure of the group and to provide a robust typing scheme.

Ideally, a typing scheme should depend on criteria that can be reproduced satisfactorily in other laboratories all over the world and if possible, a mechanism for the addition of new strains should be built in to allow the database to expand indefinitely. At the present time, the best method for such a global typing scheme is MLST using DNA fragments from seven genes. Accordingly, an MLST scheme was developed, using 146 diverse strains from the *B. cereus* group.

Chapter 2 MATERIALS and METHODS

2.1 Strains

Bacillus anthracis (DNA only), *Bacillus cereus, Bacillus mycoides, Bacillus thuringiensis* and *Bacillus weihenstephanensis* strains were obtained from various collections, representing a wide variety of sources, (soil, sea, and food). Full details about the origins of the strains are given in Appendix A.

2.2 Growth conditions

Bacillus strains were usually grown on nutrient agar (NA) or in nutrient broth (NB) at 30°C unless otherwise indicated.

Stock cultures were preserved as spore and cell suspensions in 20% v/v glycerol in water at -20°C.

2.3 DNA extractions

DNA was extracted using the 'Puregene DNA Isolation Kit', from Gentra. The manufacturer's method was adapted to give optimum yield from *Bacillus* strains. The volume of cells taken and the amount of lytic enzyme used was increased, and the final volume for re-suspension decreased, as follows: -

A loop full of culture from a nutrient agar plate was inoculated into 10 ml of Nutrient Broth with 0.5 % glucose and grown overnight at 30°C overnight, with shaking, to late exponential phase. 1-2 ml of cell suspension was added to a 1.5 ml microfuge tube, and centrifuged for 1 min at 12,000 rpm (10,000 x g). The supernatant was removed using a pipette. 300 µl of cell suspension solution was added to the cell pellet, and pipetted up and down to resuspend the cells. 2.5 µl of lytic enzyme solution was added and the tube inverted 25 times to mix thoroughly. The tubes were incubated at 37°C for 1 hour to digest the cell walls and mixed occasionally during this time. Samples were centrifuged at 12,000 rpm (10,000 x g) for 1 min to pellet the cells and the supernatant was removed. 300 µl of cell lysis solution was added to the pellet and pipetted up and down to lyse the cells. Samples were heated at 80°C for 5 min to ensure complete lysis.1.5 µl

of RNAse solution was added to the cell lysate and the sample inverted 25 times to mix, followed by incubation for 1 h at 37°C. 100 µl of protein precipitation solution was added to the cell lysate and vortexed vigorously for 20 s. Samples were placed on ice for 1 hour and then centrifuged at 12,000 rpm (10,000x g) for 3 min. The supernatant containing the DNA was poured into a new 1.5 ml microfuge tube containing 300 µl of propan-2-ol. The sample was mixed by inverting 50 times and centrifuged at 12,000 rpm (10,000 x g) for 1 min. At this stage, the DNA was visible as a small white pellet. The supernatant was poured off and the tube inverted on clean absorbent paper to drain. 300 µl of 70 % ethanol was added and the tube inverted several times to wash the pellet. The sample was centrifuged at 12,000 rpm (10000 x g) for 2 min and the ethanol carefully removed. The tube was inverted on clean paper and left to dry for 15 to 30 min. 60 µl of DNA hydration solution was added to the tube and the sample left overnight at room temperature to rehydrate. Samples were then stored at -20°C.

2.4 Multilocus Sequence Typing (MLST)

12 samples of *B. anthracis* DNA (supplied by L. Baillie, Porton Down), 58 strains of *B. cereus*, 3 strains of *B. mycoides*, 1 strain *of B. pseudomycoides*, 71 strains of *Bacillus thuringiensis* and 2 strains of *B. weihenstephanensis* were examined at seven gene loci. Details of the strains used for MLST are given in Table 3.1.

2.4.1 Primers used in MLST

When primers were designed for this project, the *Bacillus anthracis* genome had not been sequenced. However, the unfinished sequence data were available at www.tigr.org. Four genes from *S. aureus* that were used in MLST (Enright *et al.*, 2000) were selected and the equivalent gene found by a BLAST search in the unfinished *B. anthracis* genome. Primers were designed to amplify the homologous allele in *B. anthracis*. The remaining three genes were from *B. cereus* (Økstad *et al.*, 1999). Primers were tested with a selection of *B. cereus* and *B. thuringiensis* strains, to check that the correct allele would amplify.

Table 2.1 Genetic loci analysed in MLST, based on the *B. anthracis* Ames genome

Gene	Protein	*Accession no.*	Genomic position	Fragment length (bp)	Total length of gene (bp)
glpF	Glycerol uptake facilitator protein.	*BA1025*	1014815	381	822
gmk	Guanylate kinase, putative.	*BA4009*	3688226	504	618
ilvD	Dihdroxy-acid dehydratase.	*BA1853*	1736221	393	1674
pta	Phosphate acetyltransferase.	*BA5636*	5122669	414	972
pur	Phosphoribosylamino imidazolecarboxamide formyltransferase.	*BA0298*	306074	348	1,536
pycA	Pyruvate carboxylase.	*BA4157*	3809749	363	3,447
tpi	Triosephosphate isomerase.	*BA5366*	4861379	435	756

The genes were reasonably well distributed around the chromosome of *B. anthracis*, as shown in Figure 3.1 (Chapter 3 *Results*).

Table 2.2 Sequences and Annealing temperatures of primers used in MLST

Gene	Forward Primer (5' to 3')	Reverse Primer (5' to 3')	Annealing Temp.°C
glpF	GCG TTT GTG CTG GTG TAA GT	CTG CAA TCG GAA GGA AGA AG	**59**
gmk	ATT TAA GTG AGG AAG GGT AGG	GCA ATG TTC ACC AAC CAC AA	**56**
gmk2 *	ATC GTT CTT TCA GGA CCT TC		**56**
ilvD	CGG GGC AAA CAT TAA GAG AA	GGT TCT GGT CGT TTC CAT TC	**58**
ilvD 2 **	AGA TCG TAT TAC TGC TAC GG	GTT ACC ATT TGT GCA TAA CGC	**58**
pta	GCA GAG CGT TTA GCA AAA GAA	TGC AAT GCG AGT TGC TTC TA	**58**
purH	CTG CTG CGA AAA ATC ACA AA	CTC ACG ATT CGC TGC AAT AA	**56**
pycA	GCG TTA GGT GGA AAC GAA AG	CGC GTC CAA GTT TAT GGA AT	**57**
tpi	GCC CAG TAG CAC TTA GCG AC	CCG AAA CCG TCA AGA ATG AT	**58**

* An alternative forward primer for *gmk* that was sometimes necessary.

** For emetic toxin producing strains of *B. cereus*, an alternative primer set was used.

2.4.2 Polymerase Chain Reaction amplifications for MLST

The polymerase chain reaction conditions for the MLST genes were as follows: -

Reagent	**Quantity**
Taq-dNTPs buffer mix	10.0 µl
Primer (Forward)	12.5 pmol
Primer (Reverse)	12.5 pmol
$MgCl_2$ (50mM)	2.0 µl
DNA template	0.5 -1 µg

Sterile millipure water was added to make a final volume of 50 µl. 0.3 µl. (3 units) of Taq DNA polymerase (Bioline) was added after an initial denaturation at 95 °C for 5 min. The PCR conditions were 30 cycles of denaturation at 95°C for 1 min, annealing at

x°C for 1 min, extension at 72°C for 2 min, followed by a final extension at 72°C for 10 min and a completion holding temperature of 4°C.

5 µl of the product from the PCR was mixed with 2.5 µl of gel - loading buffer and electrophoresed in a 1% agarose gel, prepared in 1x TAE, containing 1 µl of ethidium bromide (10 mg ml^{-1}) in 100 ml of agarose solution. Products were viewed under UV light and photographed using a digital camera.

2.4.3 Polyethylene glycol purification (PEG)

PEG purification of PCR products was carried out to remove any unreacted primers and/or left over dNTPs, giving a very clean trace-file chromatogram (Embley, 1991).

The PCR products (50 µl) were transferred from PCR tubes to 1.5 ml microfuge tubes. 60 µl of a solution of 20 % PEG_{8000} in 2.5 M NaCl was added to each tube and mixed. The tubes were incubated overnight at 4ºC and the precipitated DNA pelleted by centrifugation at 12,000 rpm (10,000 x g) for 20 min. The supernatant was removed with a pipette and 0.5 ml of 70% cold ethanol added, followed by centrifugation at 12,000 rpm (10,000 x g) for 10 min. As before, the supernatant was removed with a pipette and the sample centrifuged at 12,000 rpm (10,000 x g) for 10 min. The supernatant was removed and the precipitated DNA dried in a heating block at 80º C for 1 to 2 min until dry. The pellets were resuspended in 5 to 10 µl (depending on the brightness of the band on the gel) of sterile millipure water. 1 to 2 µl of purified product was taken used as template in the cycle sequencing reaction.

2.4.4 Cycle sequencing reactions of PCR products

In the cycle sequencing reaction, ABI Prism BigDye terminators (dRhodamine acceptor dye linked to a fluorescein donor dye) were added to the sample DNA.

The Terminator Ready Reaction Mix is made up of the following:-

A-Dye Terminator labelled with dichloro (R6G)
C-Dye Terminator labelled with dichloro (ROX)

G-Dye Terminator labelled with dichloro (R110)
T-Dye Terminator labelled with dichloro (TAMRA)
Deoxynucleoside triphosphates (dATP, dCTP, dITP, dUTP)
AmpliTaq DNA Polymerase
$MgCl_2$
Tris-HCl buffer, pH 9.0.

Cycle sequencing was carried out using the same primers as were used in the PCR reactions.

Reaction mixtures were as follows: -

Reagent	**Quantity**
Terminator Ready Reaction Mix	2.0 µl
2.5 x Sequencing buffer	6.0 µl
Template	100 - 200 ng
Primer (F or R)	3.2 pmol
Sterile millipure water	As required to make 20 µl total volume

The cycle sequencing reaction conditions were: an initial hold at 95°C for 1 min, followed by 25 cycles of 96 °C for 10 s, 50 °C for 5 s, 60 °C for 4 min, ending with a final holding temperature of 4 °C.

2.4.5 Purification of products after cycle sequencing

Initially, the products from cycle sequencing were precipitated using isopropanol. While this worked very well, there were high peaks at the start of the sequence, and the first 20 to 30 bases were not resolved. Later samples were purified using a mixture of ethanol, sodium acetate and water, and this method gave cleaner sequence data.

Isopropanol method

The products from the cycle sequencing reaction were transferred to 1.5 ml microfuge tubes. 20 µl of sterile millipure water and 60 µl of 100% isopropanol were added, the tubes were closed and vortexed briefly. The tubes were left at room temperature for 5

min to precipitate the products and microfuged for 20 min at 10,000 x g. The supernatants were removed using a separate pipette for each sample. 250 µl of 75% isopropanol was added to the tubes and vortexed briefly. The tubes were replaced in the microcentrifuge in the same orientation as before and centrifuged for 5 min at 12,000 rpm (10,000 x g). The supernatants were again removed using a pipette and the samples were dried in a heating block at 90°C for about 1 min, until dry.

Ethanol/sodium acetate /water method

The products from the cycle sequencing reaction were transferred to 1.5 ml microfuge tubes. A mixture of the following reagents was added to each tube: - 62.5 µl of 95% ethanol, 3.0 µl of 3M sodium acetate and 14.5 µl sterile millipure water. The final volume was 80 µl per sample. The tubes were closed and vortexed briefly. The tubes were left at room temperature for 5 min to precipitate the products. The tubes were microfuged for 20 min at 12,000 rpm (10,000 x g). The supernatants were removed using a separate pipette for each sample. 250 µl of 75% ethanol was added to the tubes and vortexed briefly. The tubes were replaced in the microcentrifuge in the same orientation as before and centrifuged for 5 min at 12,000 rpm (10,000 x g). The supernatants were again removed using a pipette and the samples were dried in a heating block at 90°C for about 1 min, until dry.

Template suppression reagent (25 µl) was added to the dried DNA, vortexed briefly and centrifuged. Samples were denatured by heating for 2 min at 95°C, followed by vortexing and centrifugation. Samples were chilled on ice, transferred to sequencing tubes with septa and were kept at -20 °C until ready to be loaded onto the sequencer.

2.5 DNA Sequencing

Sequencing of the gene fragments obtained from PCR were carried out in an ABI 310 single capillary instrument at Heriot-Watt University, or in an ABI 3700 96-capillary sequencer at Oxford University.

2.5.1 Processing of sequence data

All alleles were sequenced in both directions. Sequences were assembled and edited using the 'STADEN' package (GAP4). The consensus sequence was saved and compared with existing alleles. If there were any base changes at all, the sequence was given a new allele number. Allele numbers were added to the MLST database at Oxford University and eventually each completed strain had 7 allele numbers, forming its allelic profile, or ST (Sequence Type).

2.5.2 Analysis of sequence data

The sequence data were aligned for each allele and for the concatenated sequence of 2838 bp. Neighbour joining trees were constructed using ClustalW and visualised using Treeview.

Several computer programs were used in the analysis of the sequence data: -

'ClustalW' (Thompson, J, D. Higgins, D, G. and Gibson, T, J.1994)
Available on-line at the web site at the European Bioinformatics Institute .
(http://www.ebi.ac.uk/clustalw/)
This program performs multiple alignments on nucleotide or amino acid sequences. This step was performed prior to tree construction using ClustalW. A multiple alignment is built up heuristically by a series of pairwise alignments. The most closely related sequences are aligned first and the more distant ones added. ClustalW uses the neighbour joining method (NJ) (Saitou, N. and Nei, M. (1987) for tree construction. In addition to producing a multiple sequence alignment, the ClustalW web site at 'EBI' also provides a tree file, with a .dnd extension, which can be opened by 'Treeview'.

Bootstrap analysis of the NJ trees was not available on-line and was carried out using the program ClustalX .

'Treeview' version 1.6.6 (R. Page , Glasgow University. 2001)
(http://taxonomy.zoology.gla.ac.uk/rod/treeview.html)
Treeview enables the results of the multiple sequence alignment to be viewed as a phylogenetic tree and can be represented unrooted or rooted. The user can decide which strain is to be considered the 'root'. Trees visualised by Treeview can be saved as '.wmf 'or '.emf' files and inserted into 'Powerpoint', or 'Word', to be prepared for presentation.

'START' (Sequence Typing and Recombinational Tests) (Jolley, K.A., Feil, E. J., Chan M-S., and Maiden, M, C, J. 2001). Available from the MLST web-site: -
(http://pubmlst.org/software/analysis/start/).
The 'START' program includes techniques for data summary, polymorphisms, G + C content, lineage assignment, estimation of recombination and selection. START requires two types of data input. The sequences of the alleles, in FASTA format and the allelic profile which can be pasted from an excel file, or entered directly from the keyboard. All strains may be processed for analysis, or a sub-set can be selected. START can create a 'distance matrix' using only the allelic profile data. Distances are calculated for each pair of isolates. Identical isolates have a distance of 0, and those with no loci in common have a distance of 1. START also has a useful facility for converting fasta files of the DNA sequences to nexus format for split decomposition analysis using SplitsTree.

'SplitsTree' version 3.2 (Huson, D. H., University of Bielefeld, Germany, 1997)
(http://bibiserv.techfak.uni-bielefeld.de/splits/)
SplitsTree uses the method of split decomposition to visualise evolutionary data. The program requires sequence data to be in nexus format. Using SplitsTree, ideal data will produce a tree - like structure, but other data will give rise to a network, which may be evidence for conflicting relationships. If strains have undergone recombination, the best representation of the data may be a network, rather than a bifurcating tree. The SplitsTree program can be downloaded and the splits graphs saved as 'postscript' files, which can be opened in 'Photoshop' for editing.

'DnaSP' DNA Sequence Polymorphism (version 3.5) (Rozas, J., & Rozas, R., 1998) (http://www.ub.edu/dnasp/)

DnaSp is a program for the analysis of DNA polymorphism and divergence. DnaSP compares one population with a second population, and accepts data in fasta, mega, nexus and other formats. Tests include Fst (gene flow), numbers of fixed differences and shared mutations.

'PDBsum' (Laskowski, R. European Bioinformatics Institute)
http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/

Protein Data Base is a pictorial database providing an overview of every macromolecular structure deposited in the Protein Data Bank (PDB). It provides schematic diagrams of the molecules in each structure and of the interactions between them. Entries are accessed by their PDB code or by pasting in a new sequence.

'SAS' (Laskowski, R. European Bioinformatics Institute)
http://www.ebi.ac.uk/thornton-srv/databases/sas/

Secondary Annotated Structure database, provides predicted structure diagrams from amino acid sequences and is accessible from a link to PDB.

2.6 Random Amplified Polymorphic DNA (RAPDs)

PCR using a single short primer to produce patterns of multiple bands on an agarose gel was carried out using the primer described by Brousseau *et al.,* (1993), 0955-03, 5'-CCG AGT CCA-3'.

The PCR conditions, for the RAPDs were as follows:-

Reagent	**Quantity**
Taq-dNTPs buffer mix	10.0 µl
Primer (100pmol)	3.0 µl
$MgCl_2$ (50mM)	5.0 µl
DNA template	0.5 - 1 µg

Sterile millipure water was added to make a total volume of 50 µl.

The temperature profile for the application was 5 min at 96°C, addition of 0.3 µl (3 units) of Taq polymerase, followed by 30 cycles of 96°C for 10 s, 30°C for 10 s and 72°C for 10 s. A final extension step for 7 min at 72°C was carried out and the samples were held at 4°C until electrophoresis could be carried out.

Electrophoresis was performed in a 1.5 % agarose gel in 0.5 x Tris borate buffer (TBE), with 1µl of ethidium bromide (10mg ml^{-1}) in 100 ml of agarose, for 2.5 hours at 60 v/cm, The results were visualised under a UV light and recorded using a digital camera.

2.7 Crystal toxin genes

Selected strains of *Bacillus thuringiensis* were investigated for the presence of several crystal toxin genes, using published primers.

Table 2.3 Primers for detection of crystal toxin genes in various *B. thuringiensis* serotypes.

Crystal toxin gene	Primer	Sequence 5' – 3'	Anneal temp. °C	Product size bp	Ref.
Cry 1	gral/cry1/d	CTGGATTTACAGGTGGGGATAT	52	558	Bravo *et al.*, 1998
	gral/cry1/r	TGAGTCGCTTCGCATATTTGACT			
Cry 1B	CJ8	CTTCATCACGATGGAGTAA	53	367	Ceron *et al.*, 1994
	CJ9	CATAATTTGGTCGTTCTGTT			
Cry 1E	1E	TAGGGATAAATGTAGTACAG	45	1137	Juarez-Perez *et al.*, 1997
	1(-)	MDATYTCTAKRTCTTGACTA			
Cry 3A	Un3(d)	CGTTATCGCAGAGAGATGACATTA	56	951	Ben-Dov *et al.*, 1997
	EE-3Aa (r)	TGGTGCCCCGTCTAAACTGAGTGT			
Cry 4	Dip1A_for	CAAGCCGCAAATCTTGTGGA	55	797	Carozzi *et al.*, 1991
	Dip1B_rev	ATGGCTTGTTTCGCTACATC			
Cry 5,12,14, 21 Non-specific	gral-nem (d)	TTACGTAAATTGGTCAATCAAGCA	62	474-489	Bravo *et al.*, 1998
	gral-nem (r)	AAGACCAAATTCAATACCAGG GTT			

2.8 Sodium Dodecyl Sulphate - Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Two methods of preparing samples for SDS-PAGE were carried out.

2.8.1 SDS-PAGE (modified from Park *et al.*, 1998)

A loop full of culture from a NA plate was inoculated into 100 ml NYSM (Nutrient agar, yeast and salts) or NB broth (no glucose) and grown until sporulation (usually about 3-4 d), at 30ºC with shaking.

50ml aliquots of culture were centrifuged at 6,000 rpm (3,000 x g) for 20 min, recombined and the pellet resuspended in 1 ml of sterile millipure water, forming a very thick suspension of cells and crystal protein. Sample (100 µl) was added to 100 µl of 2 x sample buffer, in 1.5 ml microfuge tubes and boiled for 10 min. Aliquots of 5 µl and 10 µl were electrophoresed in a 9 % separating gel. Standard marker proteins (5 µl of known molecular weight were included (Sigma - High Molecular Weight Standard Mixture, stock no SDS-6H).

2.8.2 SDS-PAGE (modified from Lecadet *et al.*, 1992)

A loop full of culture from a NA plate was inoculated into 100 ml NYSM broth (no glucose) and grown until sporulation (usually about 3-4 d), at 30º C with shaking.

1.5 ml of growth culture was transferred to a 1.5 ml microfuge tube and centrifuged at 12,000 rpm (10,000 x g) for 10 min. The supernatant was discarded and the pellet resuspended in 1.5 ml of 0.5 M NaCl. Samples were washed twice by centrifugation and resuspended in 0.75 ml of cold, sterile millipure water containing 1 mM of phenylmethylsulfonyl fluoride (PMSF is a protease inhibitor). Samples were resuspended in 375 µl of 1 mM aqueous PMSF and stored at -20º C. Sample (25 µl) was added to 25 µl of 2 x sample buffer in a 1.5 ml microfuge, denatured by heating at 100º C for 10 min in a heating block and 5µl samples were electrophoresed on a 9% separating gel.

9% Separating Gel	
40 % Acrylamide	8 ml
2 % BisAcrylamide	5.5 ml
4 x Resolving gel Buffer	9.25 ml
Distilled water	12.3 ml
10 % Ammonium Persulphate	0.2 ml
TEMED	20 µl

The separating gel was overlain with a 4.5 % stacking gel.

4.5% Stacking Gel	
40 % Acrylamide	1.1ml
2 % BisAcrylamide	0.7 ml
4 x Stacking gel Buffer	2.5 ml
Distilled water	5.6 ml
10 % Ammonium Persulphate	30 µl
TEMED	10 µl

Electrophoresis was carried out using a mini-PROTEAN II Electrophoresis Cell from Bio-Rad. Running buffer was added to the upper buffer chamber of the gel tank to mid-way between the short and long plates. Buffer was added to the lower buffer chamber, to cover 1 cm of the gel. Sample (5 µl) was added to each well and electrophoresis was conducted at 200 volts for about 45 min.

The proteins were stained with coomassie blue (0.1%) for one hour and then destained.

Stain for SDS-Page gels	
Coomassie blue (R250)	0.5 g
Methanol	250 ml
Glacial acetic acid	50 ml
Distilled water	200 ml

Destain for SDS-Page gels	
Methanol	100 ml
Glacial acetic acid	100 ml
Distilled water	800 ml

2.9 Western blot

Following SDS-PAGE, the protein on the gel was transferred (blotted) onto a membrane, using a Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad, catalogue numbers 170-3930 and 170-3935). The membrane used was optimised nylon (0.45 µM Electran BDH). After transfer, the membrane was probed with antibody to cry1. The antibody, raised in a rabbit, was kindly provided by Dr Vincent Sanchis, Institut Pasteur: - antiserum anti-crystals of a *B. thuringiensis* strain serotype 6 entomocidus (This strain produces Cry 1A and Cry 1C toxins).

The nitrocellulose membrane, plus blotting paper were soaked in transfer buffer (25 mM Tris, 192 mM glycine and 20% v/v methanol, pH 8.3) for about 5 min before being assembled into the transfer cassette, as follows: -

One Scotch-Brite pad was placed onto the transfer cassette and a sheet of pre-soaked blotting paper placed on top of it. The gel was placed on top of the blotting paper and bubbles carefully removed by rolling a glass pipette over the surface. An orientation mark was made on the membrane and the membrane positioned on top of the gel. Bubbles were again removed and a second sheet of pre-soaked blotting paper placed on top of the membrane. Bubbles were removed and a second Scotch-Brite pad placed on top of the assemblage. The cassette was closed and positioned in the transfer cell. A cooling unit, containing ice was included in the transfer container to prevent overheating. Protein was transferred from the gel to the membrane for 1 h at 100 volts.

Following transfer, the gel was stained with coomassie blue (0.1%), to check the efficiency of transfer, (there should not be any protein left on the gel). The membrane was placed in TTBS + 10% milk for 1 h, or overnight. 5 ml of the first antibody solution *(anti Cry1) was added, and incubated for 1 h at room temperature, with gentle shaking.

The membrane was washed three times (15 min for each wash) with 10 ml TTBS. 5 ml of the second antibody solution *(anti-rabbit IgG-Alkaline phosphate solution) was added and incubated for 1 h at room temperature with gentle shaking. The membrane was again washed three times (15 min for each wash) with 10 ml TTBS. Ten ml of colour development solution was added to the membrane and incubated in the dark. The membrane was inspected after a few minutes to check for colour development, which should be complete after about 20 minutes. Finally, the membrane was rinsed with distilled water to stop the reaction and dried on filter paper before being photographed with a digital camera.

*Antibody dilutions

Various dilutions of antibody were tried to optimise the result, starting with 1:1,000 for both 1^{st} and 2^{nd} antibodies. This was found to be too strong. The final dilutions were 1: 20,000 for both antibodies.

2.10 Electron microscopy

Strains for examination by electron microscopy were grown in 100 ml NSM (Nutrient broth + salts) at 30°C, with shaking until sporulation and lysis of most of the cells (about 3 to 4 d). Culture (10 ml) was centrifuged at 3,000 x g for 20 min, washed twice in distilled water and resuspended in 5 ml of distilled water. 50 µl of sample was pipetted onto an EM stubb (covered with a tab) and allowed to dry overnight.

Samples of spores and crystals were observed using a Philips XL30 environmental scanning electron microscope. The best definition was obtained when the samples were coated with gold before examination. Electron microscopy was carried out at the ESEM facility at the Institute for Petroleum Engineering, managed by Dr J. Buckman.

2.11 Mosquitocidal activity of *Bt* israelensis

Several strains of *Bacillus thuringiensis* were assessed for mosquitocidal activity. The main purpose of this work was to determine if a *Bt* israelensis strain with an unusual ST was pathogenic to mosquito larvae. Strains from other serotypes were included as negative controls which were not expected to show pathogenicity to mosquito larvae. A typical *Bt* israelensis strain was included as a positive control.

Strains to be tested were grown for three days in NB + salts and serial dilutions made to the culture as shown in Table 2.4. The tests were carried out in square plastic trays ($10cm^2$) that were divided into a grid of 25 small cells, each cell being $2cm^3$.

Tests were carried out in duplicate: -

Table 2.4 Dilutions for mosquitocidal assay

cell	1	2	3	4	5
dilution	1 ml from culture + 1ml water	1 ml from cell 1 + 1 ml water	1 ml from cell 2 + 1 ml water	1 ml from cell 3 + 1 ml water	1 ml from cell 4 + 1 ml water

Five healthy, active mosquito larvae were placed in each cell, and observed for activity at intervals from one hour to 24 hours.

2.12 Starch hydrolysis

Starch-agar plates were made up with NA plus 1% starch and sterilised at 115°C for 10 min. A loopfull of culture from a standard NA plate was streaked lightly once over the starch-agar plate and incubated at 30°C for 4 days.

Lugol's iodine (5x), 5.0 g of iodine, 10.0 g of potassium iodide and 100 ml distilled water, was made up and diluted 1:5 with distilled water before use. After incubation, the plate was flooded with the diluted iodine solution. Unchanged starch turns blue-black, hydrolysis of starch is indicated by a clear zone around the streak. Weakly positive

strains may not produce a distinct clear zone, around themselves and the culture was scraped away to reveal a clear area underneath the culture.

2.13 Enterotoxin Genes

Bacillus cereus strains were tested for the presence of two enterotoxin genes. Established primers were used to detect the haemolytic enterotoxin, *hbl* and the non-haemolytic enterotoxin *nhe* genes. New primers were designed to detect the occurrence of genes within the *hbl* and the *nhe* operons. Details of the primers are shown in Table 2.5 and the positions of these primers are shown in Figure 2.1.

Table 2.5 Primers for amplification of enterotoxin genes in *B. cereus* group.

Primer	Sequence (5' - 3')	Annealing temperature (°C)	Size of product (base-pairs)
nhe A1F	ATG TAT CGT CTG TTG ATG CG	51	1200
nhe B1R	TAC TTG ATG TCG TTT GTG CC		
nhe A1301F	GTT GAA GAA GGC ACG TAT AC	55	1716
nhe C3017R	CAG CTG GTC CAA GAG AAT AG		
hbl CF1	GGA ATG GTC ATC GGA ACT CT	53	1900
hbl DR1	CTG CTT GAT TAG CAC GAT CT		
hbl C11729F	CCG CAG CAA AAA CAC AAG AG	55	1892
hbl A13621R	CTC CAG CCT TTT GCA AGG TC		

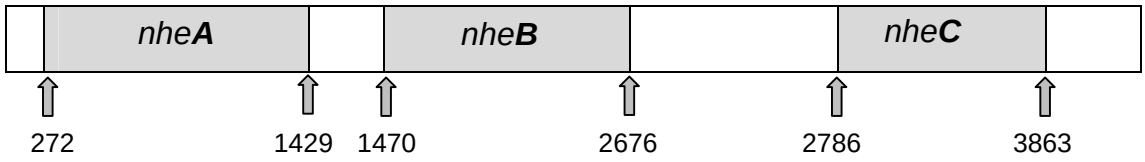
A. *nhe* operon *Bacillus cereus* strain 1230-88 isolated from a food poisoning outbreak in Norway (Accession Y19005; Granum *et al.*,1999)

***nhe* primers for genes B to C**

Primer nheA1F starts at 1741 bases and primer nheB1R starts at 2926 bases.

New *nhe* primers for genes A to C

Primer NheA1301F starts at 1301 bases and primer NheC3017R starts at 3017 bases.



B. *hbl* operon Chromosomal locus of ATCC 14579 (Accession AJ237785; Økstad *et al.*, 1999

***hbl* primers for genes C to D**

Primer hblCF1 starts at 11088 bases and primer hblDR1 starts at 13143 bases.

New *hbl* primers for genes C to A

Primer hblC11729F starts at 11729 bases and primer hblA13621R starts at 13621 bases.

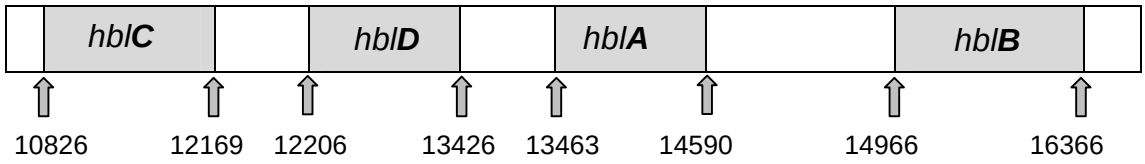


Figure 2.1 *Bacillus cereus* toxin genes and positions of primers: - A. *nhe,* B. *hbl*

Chapter 3 RESULTS

3.1 Multilocus sequence typing (MLST)

3.1.1 Strain collection

The *Bacillus cereus* group is very diverse and strains were assembled to represent as much of the known variation as possible. Strains of *B. cereus* from milk, other foods, packaging materials, emetic food poisoning outbreaks and environmental isolates were obtained. Some strains (S numbers) were already present in the laboratory collection, having been part of an earlier numerical taxonomic study of the Genus *Bacillus* (Priest *et al.,* 1988).

Strains were donated, on request, from other culture collections, in particular, the collections of M. Aquino de Muro, A-B Kolstø, S. Scherer and M. S. Salkinoja-Salonen. Strains representing numerous serotypes of *B. thuringiensis* were donated by the Institut Pasteur. Later in the project, some *B. cereus* strains isolated from clinical infections were provided by J. McLaughlin (Health Protection Agency), and B. Thakker (Glasgow Royal Infirmary). It was not possible to include strains of *B. anthracis,* but DNA preparations from 12 representative strains of *B. anthracis* were supplied by L. Baillie (Priest *et al.,* 2004). A total of 146 strains was analysed using the technique of MLST, and these are listed in Table 3.1, in order of sequence type (ST).

Full details of the collection of strains, including source, country of origin and date of isolation are included in Appendix A.

3.1.2 Allele selection

3.1.2 i Choice of alleles

The genes selected for MLST should be essential 'housekeeping' genes, which are present in all strains, and not subject to unusually fast rates of evolution. Genes coding for antibiotic resistance, sporulation or external structures such as capsules or flagellae, are unsuitable for use in a primary MLST scheme because of their likely uneven rates of evolutionary change. Such genes may be useful later, when investigating particular clonal groups or lineages (Cooper & Feil, 2004).

Four loci, *glpF*, *gmk*, *pta* and *tpi* were adapted from the MLST analysis of *S. aureus*, a related low G + C Gram positive bacterium (Enright *et al.*, 2000). The nucleotide sequences of these genes were obtained from the *B. anthracis* genome sequence by BLAST searches at TIGR (www. tigr.org). The other three loci, *ilvD*, *pur* and *pycA* were derived from sequences described by Økstad *et al.*, (1999),

The seven genes selected were reasonably well distributed around the chromosome of *B. anthracis* Ames, as shown in Figure 3.1.

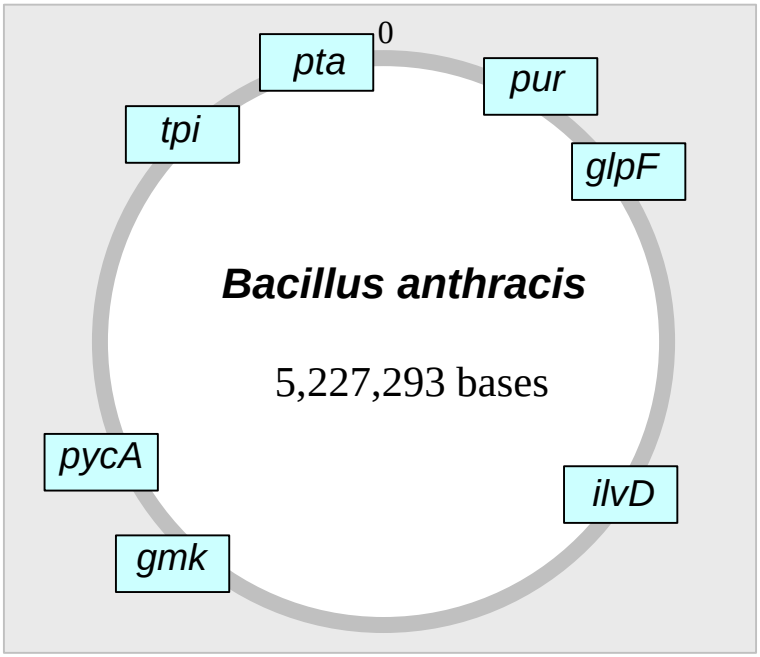


Figure 3.1 Position of alleles in *B. anthracis* chromosome

3.1.2.ii Protein function

The function of the proteins coded for by the MLST genes was investigated using on-line databases available through PubMed and EBI with links to Prosite, Pfam and PDB. The sequences for the complete genes were derived from the whole genome of the Ames strain of *B. anthracis* (Read *et al.,* 2003) and used in BLAST searches:-

GlpF *BA1025* (glycerol uptake facilitator protein) is a membrane protein and catalyses the transport of glycerol and other small molecules by forming a transmembrane channel. X-ray studies on glpF from *E. coli* have shown an alpha-barrel structure, surrounded by six transmembrane helices and 2 half-spanning helices (Fu *et al.,* 2002).

Gmk *BA4009* (guanylate kinase) catalyses the following reaction (signal transduction system): -
ATP (adenosine triphosphate) +GMP (guanosine monophosphate) ⇔ ADP + GDP

IlvD *BA1853* (dihydroxyacid dehydratase) is involved in biosynthesis of the amino acids isoleucine, leucine and valine.

Pta *BA5636* (phosphate acetyltransferase) catalyses the following reaction:-
Acetyl-CoA + phosphate ⇔ CoA + acetyl phosphate.

Pur *BA0298* (phosphoribosylaminoimidazolecarboxamide formyltransferase)
This is a member of a family of bifunctional enzymes catalysing the last two steps in de novo purine biosynthesis.

PycA *BA4157* (pyruvate carboxylase) catalyses the following reaction:-
ATP + pyruvate +HCO(3)(-) ⇔ ADP + phosphate + oxaloacetate.
PycA is an anaplerotic enzyme, replenishing oxaloacetate for the tricarboxylic acid cycle. Oxaloacetate is used catalytically in the TCA cycle, but can also be drawn off for the biosynthesis of amino acids (Stryer, 1995 b).

Tpi *BA5366* (triosephosphate isomerase) catalyzes the reversible interconversion of glyceraldehyde 3-phosphate and dihydroxyacetone phosphate. Triosephosphate isomerase has an essential role in the metabolic pathways of energy production. The enzyme is a dimer of identical subunits and the sequence around the active site is highly conserved.

3.1.2.iii Determination of primers / sequences

Primers for PCR amplification of gene fragments of around 500 bp were designed. The primers for all genes were tested on a small sample of diverse strains, and the seven most reliable and reproducible sets of primers and alleles were adopted. However, after sequencing a considerable number of strains, it became apparent that some PCR products were not being amplified from some strains. In particular, the emetic toxin producing strains of *B. cereus* did not amplify a PCR product with the initial *ilvD* primers, and a second pair of primers was designed for those strains. Interestingly, the new *ilvD2* primers did not work with other *B. cereus* strains, or with *B. thuringiensis* strains. Some examples of the PCR products obtained are shown in Figures 3.2 to 3.4.
Full details of the final sets of primers are given Tables 2.1 and 2.2 (Materials and Methods).

3.1.2.iv DNA sequencing

Gene fragments of around 400 to 600 bp were obtained by PCR from each gene (e.g. Figures 3.2 to 3.4). It was important to obtain a clean single band from the PCR reaction. The PCR products were purified by PEG precipitation (Materials and Methods, 2.4.3) to remove any unreacted primers and dNTPs from the solution. The PCR products were labelled with fluorescent dyes by cycle sequencing in a thermal cycler, ethanol precipitated to re-purify, and the sequence determined in both forward and reverse directions in an ABI capillary electrophoresis DNA sequencer. Forward and reverse sequences were assembled with STADEN software (Staden, 1996) to provide unambiguous sequence over the required length of the allele

3.1.2.v Determination of sequence types

The DNA sequences for each allele were compared. Alleles that differed by one or more nucleotides were allocated arbitrary numbers, as shown in Table 3.1. The seven allele numbers formed an allelic profile and were allocated a unique ST (Sequence Type) number. In this way, each strain could be defined by a series of numbers, for example, strain S57 = (8,8,16,13,2,16,7) ST-8. The number of strains studied was 146, and a total of 88 different STs were recovered.

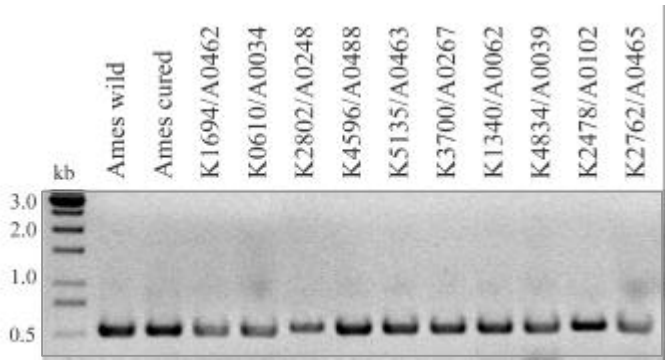


Figure 3.2 PCR amplification of *glpF* gene fragments from strains of *B. anthracis.*

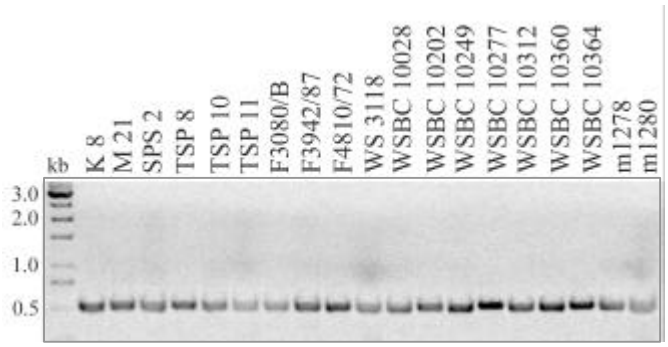


Figure 3.3 PCR amplification of *tpi* gene fragments from strains of *B. cereus , B. mycoides* (WSBC 10277, WSBC 10360), *B. pseudomycoides* (WS 3118) and *B. weihenstephanensis* (WSBC 10364).

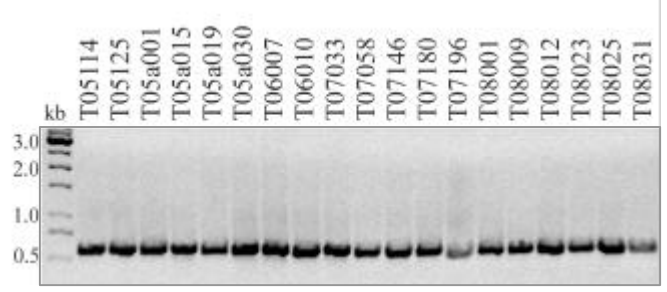


Figure 3.4 PCR amplification of *pta* gene fragments from strains of *B. thuringiensis.*

Table 3.1 Allele numbers and sequence types assigned to strains.

Strain	***species* / serovar**	**ST**	**allele**						
			glpF	***gmk***	***ilvD***	***pta***	***pur***	***pycA***	***tpi***
Ames wild type	*anthracis*	**1**	1	1	1	1	1	1	1
Ames cured_delta	*anthracis*	**1**	1	1	1	1	1	1	1
K0610/A0034	*anthracis*	**1**	1	1	1	1	1	1	1
K4834/A0039	*anthracis*	**1**	1	1	1	1	1	1	1
K1340/A0062	*anthracis*	**1**	1	1	1	1	1	1	1
K1694/A0462	*anthracis*	**1**	1	1	1	1	1	1	1
K5135/A0463	*anthracis*	**1**	1	1	1	1	1	1	1
K4596/A0488	*anthracis*	**1**	1	1	1	1	1	1	1
K3700/A0267	*anthracis*	**2**	1	1	2	1	1	2	1
K2478/A0102	*anthracis*	**3**	2	1	1	1	1	1	1
K2762/A0465	*anthracis*	**3**	2	1	1	1	1	1	1
ATCC 14579	*cereus*	**4**	13	8	8	11	11	12	7
m1545	*cereus*	**5**	6	4	3	4	14	4	3
m1564	*cereus*	**6**	4	3	4	6	13	5	6
M21	*cereus*	**7**	5	4	3	4	15	6	7
S57	*cereus*	**8**	8	8	16	13	2	16	7
S58	*cereus*	**8**	8	8	16	13	2	16	7
S59	*cereus*	**8**	8	8	16	13	2	16	7
T03a001	kurstaki	**8**	8	8	16	13	2	16	7
T03a075	kurstaki	**8**	8	8	16	13	2	16	7
T03a172	kurstaki	**8**	8	8	16	13	2	16	7
T03a287	kurstaki	**8**	8	8	16	13	2	16	7
T03a361	kurstaki	**8**	8	8	16	13	2	16	7
NCTC 6474	*cereus*	**9**	15	7	8	2	16	3	13
T01001	thuringiensis	**10**	15	6	10	8	3	7	14
T01015	thuringiensis	**10**	15	6	10	8	3	7	14
T01022	thuringiensis	**10**	15	6	10	8	3	7	14
T01326	thuringiensis	**10**	15	6	10	8	3	7	14
T01058	thuringiensis	**10**	15	6	10	8	3	7	14
2002734341	*cereus*	**11**	34	1	32	1	33	37	24
T04002	sotto	**12**	15	7	7	2	7	10	13
T04016	sotto	**12**	15	7	7	2	7	10	13
T04024	sotto	**12**	15	7	7	2	7	10	13
T15001	dakota	**12**	15	7	7	2	7	10	13
T04b001	kenyae	**13**	10	8	15	12	2	16	7
T04b012	kenyae	**13**	10	8	15	12	2	16	7
T04b054	kenyae	**13**	10	8	15	12	2	16	7
T04b060	kenyae	**13**	10	8	15	12	2	16	7
T04b073	kenyae	**13**	10	8	15	12	2	16	7

strain	***species* / serovar**	**ST**	**allele**						
			glpF	***gmk***	***ilvD***	***pta***	***pur***	***pycA***	***tpi***
T06007	entomocidus	**14**	14	8	9	7	10	15	10
T06010	entomocidus	**14**	14	8	9	7	10	15	10
T07033	aizawai	**15**	9	8	16	13	2	16	9
T07058	aizawai	**15**	9	8	16	13	2	16	9
T07180	aizawai	**15**	9	8	16	13	2	16	9
t05a030	canadensis	**16**	15	7	7	2	6	8	13
T08025	morrisoni	**16**	15	7	7	2	6	8	13
CCCT 2259	israelensis	**16**	15	7	7	2	6	8	13
T13001	pakistani	**17**	12	8	11	11	9	13	5
T13004	pakistani	**17**	12	8	11	11	9	13	5
T13028	pakistani	**18**	11	9	14	12	12	14	7
1725601_V	*cereus*	**18**	11	9	14	12	12	14	7
WSBC 10249	*cereus*	**19**	12	8	9	11	11	12	8
m1280	*cereus*	**19**	12	8	9	11	11	12	8
WSBC 10312	cereus	**20**	16	6	12	9	4	11	15
WSBC 10277	*mycoides*	**21**	18	9	5	15	17	9	11
T09010	tolworthi	**22**	12	8	8	11	9	12	23
T09011	tolworthi	**22**	12	8	8	11	9	12	23
T09024	tolworthi	**22**	12	8	8	11	9	12	23
T09034	tolworthi	**22**	12	8	8	11	9	12	23
T09052	tolworthi	**22**	12	8	8	11	9	12	23
T08001	morrisoni	**23**	15	7	7	2	5	8	13
T08009	morrisoni	**23**	15	7	7	2	5	8	13
T08012	morrisoni	**23**	15	7	7	2	5	8	13
T08023	morrisoni	**23**	15	7	7	2	5	8	13
T08031	morrisoni	**23**	15	7	7	2	5	8	13
4Btt	tenebrionis	**23**	15	7	7	2	5	8	13
NCIB 6349	*cereus*	**24**	12	8	9	14	11	12	10
Cal3	*cereus*	**24**	12	8	9	14	11	12	10
WSBC 10028	*cereus*	**24**	12	8	9	14	11	12	10
T05005	galleriae	**25**	30	8	16	13	2	16	7
T05033	galleriae	**25**	30	8	16	13	2	16	7
T05114	galleriae	**25**	30	8	16	13	2	16	7
S73	*cereus*	**26**	3	2	31	5	16	3	4
S74	*cereus*	**26**	3	2	31	5	16	3	4
S710	*cereus*	**26**	3	2	31	5	16	3	4
F3080B/87	*cereus*	**26**	3	2	31	5	16	3	4
F3942/87	*cereus*	**26**	3	2	31	5	16	3	4
F4810/72	*cereus*	**26**	3	2	31	5	16	3	4
s70	*cereus*	**27**	34	1	17	16	18	33	24
s71	*cereus*	**28**	19	2	18	17	19	3	2

strain	***species* / serovar**	**ST**	**allele**						
			glpF	***gmk***	***ilvD***	***pta***	***pur***	***pycA***	***tpi***
S86	* *cereus 'terminalis'*	**29**	20	8	8	35	8	17	17
S363	*cereus*	**30**	17	5	6	3	16	34	12
S366	*cereus*	**31**	21	2	19	5	16	18	2
ATCC 10987	*cereus*	**32**	5	4	3	4	15	6	16
ATCC 10876	*cereus*	**33**	22	9	18	18	20	19	7
ATCC 11778	*cereus*	**34**	33	8	13	19	2	17	17
AH621	*cereus*	**35**	23	10	5	15	21	23	11
AH647	*cereus*	**36**	18	11	3	20	22	22	18
AH684	*cereus*	**37**	18	10	14	20	22	22	18
ATCC 4342	*cereus*	**38**	24	12	14	21	23	31	19
SPS 2	*cereus*	**39**	37	9	18	12	24	21	7
TSP 11	*cereus*	**40**	39	9	14	12	24	14	7
WSBC 10202	*weihenstephanensis*	**41**	25	10	22	36	26	23	11
WSBC 10364	*weihenstephanensis*	**42**	26	21	12	25	27	32	18
m1278	*cereus*	**43**	27	20	23	8	4	24	21
m1292	*cereus*	**44**	28	9	14	12	2	25	7
m1293	*cereus*	**45**	19	14	21	5	19	3	2
m1550	*cereus*	**46**	33	8	13	10	8	17	7
m1576	*cereus*	**47**	29	2	24	5	32	3	22
T01246	thuringiensis	**48**	13	8	25	11	9	35	10
T04236	sotto	**49**	15	7	15	2	7	10	13
T05a001	canadensis	**50**	13	16	9	28	4	12	8
T05a015	canadensis	**51**	31	9	14	29	28	27	7
T05a019	canadensis	**52**	32	19	26	30	9	28	10
T07146	aizawai	**53**	33	17	27	31	29	17	7
T07196	aizawai	**54**	11	9	14	32	12	14	7
T10016	darmstadiensis	**55**	15	7	8	2	7	26	13
T10001	darmstadiensis	**56**	15	7	7	2	7	26	13
T10003	darmstadiensis	**56**	15	7	7	2	7	26	13
T10017	darmstadiensis	**56**	15	7	7	2	7	26	13
T10018	darmstadiensis	**56**	15	7	7	2	7	26	13
T10024	darmstadiensis	**57**	34	15	28	33	30	29	24
T15006	dakota	**58**	35	6	29	9	4	7	21
T18001	kumamotoensis	**59**	22	9	30	18	20	21	7
T18002	kumamotoensis	**59**	22	9	30	18	20	21	7
T18004	kumamotoensis	**60**	36	15	28	34	30	29	25
T05125	galleriae	**61**	15	7	7	27	7	26	13
2002734374	*cereus*	**62**	38	1	32	1	18	33	24
T08019	morrisoni	**63**	15	24	29	11	4	39	28
T08003	morrisoni	**64**	33	8	36	19	41	17	27
T14531	israelensis	**65**	15	7	14	2	6	14	13

strain	***species* / serovar**	**ST**	**allele**						
			glpF	***gmk***	***ilvD***	***pta***	***pur***	***pycA***	***tpi***
m1552	*cereus*	**66**	40	9	14	26	37	36	7
T08121	morrisoni	**67**	15	7	34	2	5	10	13
T09016	tolworthi	**68**	22	9	34	18	20	40	29
T09043	tolworthi	**68**	22	9	34	18	20	40	29
T08016	morrisoni	**69**	15	7	36	2	5	8	13
T08051	morrisoni	**70**	15	7	36	2	37	8	13
T09020	tolworthi	**71**	41	25	12	9	4	12	15
R_3238/03	*cereus*	**72**	43	26	35	40	39	41	30
R_3149/03	*cereus*	**73**	13	8	9	14	9	12	31
R_2955/03	*cereus*	**73**	13	8	9	14	9	12	31
R_3098/03	*cereus*	**74**	42	27	37	41	40	42	32
R_3039/03	*cereus*	**75**	44	1	32	1	18	33	24
191560_K	*cereus*	**76**	34	15	28	38	30	29	33
168287_M	*cereus*	**77**	3	23	38	39	38	43	34
2002734342	*cereus*	**78**	24	22	33	37	34	38	5
S79	** cereus 'fluorescens'*	**79**	7	2	14	4	2	3	15
TSP 8	*cereus*	**80**	37	9	21	5	19	21	4
S78	** cereus 'albolactis'*	**81**	3	2	19	5	36	18	2
K8	*cereus*	**82**	45	21	20	22	42	20	26
WS 3118	*pseudomycoides*	**83**	21	13	16	23	25	44	35
2002734432	*cereus*	**84**	46	9	39	29	24	45	7
2002734433	*cereus*	**85**	47	28	14	12	2	36	7
TSP 10	*cereus*	**86**	33	8	13	19	8	17	17
WSBC 10360	*mycoides*	**87**	38	18	34	24	31	30	20
STR4	*mycoides*	**88**	18	10	22	15	35	9	11

Bacillus thuringiensis strains are designated by serovar (not italicised)

* The following three strains are described by Gordon *et al.,* (1973).

Strain S78, '*B. cereus albolactis'* (ST 81) is *B. cereus* ATCC 7004

Strain S79, '*B. cereus fluorescens'* (ST 79) is *B. cereus* ATCC 13824

Strain S86, '*B. cereus terminalis'* (ST 29) is *B. cereus* NRS 1417.

Only four STs were represented more than five times in the database. These were ST-1 (eight strains of *B. anthracis*), ST-8 (three *B. cereus* and five *Bt* kurstaki strains), ST-23 (four strains of *Bt* morrisoni and one *Bt* tenebrionis) and ST-26 (five emetic toxin-producing strains of *B. cereus*). Three STs were present five times in the dataset, two STs four times, four STs three times, eight STs two times, and the remaining 67 STs occurred only once.

3.1.2.vi Analysis of alleles

The length and frequency of each allele, number of polymorphic sites, d_N/d_S ratio and % G+C content are given in Table 3.3. The numbers of different alleles varied from 28 different *gmk* alleles to 47 *glpF* alleles. Alleles also varied considerably in the degree of sequence heterogeneity they contained, for example, there were 92 polymorphic sites for the *ilvD* allele, and only 40 polymorphic sites in *tpi*. The mean mol % G+C content varied from 38.3 for *gmk* to 44.4 for *ilvD.* The value of d_N/d_S was less than 1 for all loci, indicating that strong purifying selection had eliminated most non-synonymous changes.

Table 3.2 Allele length, frequency, number of polymorphic sites and d_N/d_S ratios of alleles

Allele	Length (number of bases)	Frequency (number of alleles)	Number of polymorphic sites	d_N/d_S ratio *	Mean % GC content
glpF	381	47	57	0.1002	38.57
gmk	504	28	89	0.0225	38.33
ilvD	393	39	92	0.0163	44.39
pta	414	41	67	0.0164	40.97
pur	348	42	82	0.0092	39.05
pycA	363	45	90	0.0270	41.06
tpi	435	35	40	0.0990	43.98

*d_N/d_S ratio is the ratio of non-synonymous to synonymous mutations and was calculated using the 'START' program, available from http://pubmlst.org/bcereus/.

Non-synonymous changes, where an amino acid is changed, are usually eliminated by selection as they are lethal or deleterious. Sometimes they may be beneficial to the organism and are retained. The ratio of synonymous to non-synonymous mutations gives information about the effect of purifying selection on the gene.

Table 3.3 Allele frequencies for 146 strains.

Allele	*glpF*	*gmk*	*ilvD*	*pta*	*pur*	*pycA*	*tpi*
1	9	15	10	14	11	10	11
2	2	11	1	24	23	1	4
3	8	1	4	1	5	11	1
4	1	3	1	4	6	1	7
5	2	1	2	11	8	1	3
6	1	7	1	1	4	2	1
7	1	25	18	2	11	6	33
8	8	42	9	6	3	11	3
9	3	16	10	3	11	2	3
10	5	4	5	1	2	6	7
11	3	1	2	12	6	1	4
12	12	1	3	11	3	15	1
13	5	1	3	14	1	2	25
14	2	1	12	5	1	5	5
15	31	3	6	3	2	2	3
16	1	1	15	1	9	19	1
17	1	1	1	1	1	6	3
18	4	1	3	5	3	2	3
19	2	1	2	3	3	1	1
20	1	1	1	2	5	1	1
21	2	2	2	1	1	4	2
22	5	1	2	1	2	2	1
23	1	1	1	1	1	2	5
24	2	1	1	1	3	1	5
25	1	1	1	1	1	1	1
26	1	1	1	1	1	6	1
27	1	1	1	1	1	1	1
28	1	1	3	1	1	1	1
29	1	-	2	2	1	3	2
30	3	-	2	1	3	1	1
31	1	-	6	1	1	1	2
32	1	-	3	1	1	1	1
33	5	-	1	1	1	3	1
34	4	-	4	1	1	1	1
35	1	-	1	1	1	1	1
36	1	-	3	1	1	2	-
37	2	-	1	1	2	1	-
38	2	-	1	1	1	1	-
39	1	-	1	1	1	1	-
40	1	-	-	1	1	2	-
41	1	-	-	1	1	1	-
42	1	-	-	-	1	1	-
43	1	-	-	-	-	1	-
44	1	-	-	-	-	1	-
45	1	-	-	-	-	1	-
46	1	-	-	-	-	-	-
47	1	-	-	-	-	-	-
Total	**47**	**28**	**39**	**41**	**42**	**45**	**35**

Alleles represented eight or more times among the 146 strains are highlighted.

3.1.2.vii Amino acid alignments of frequently occurring alleles.

An examination of Table 3.3 showed that some alleles were very common, e.g. the most prevalent allele was *gmk*-8 which occurred 42 times in the dataset, *tpi*-7 was found 33 times and *glpF*-15 was present 31 times. Other alleles were rare and many occurred only once. The common alleles, occurring eight or more times in the dataset, were compared to detect non-synonymous nucleotide changes which may result in the formation of a different protein structure. Translated sequences were aligned using ClustalW and the secondary structure of the protein predicted using PDBsum. Both programs are available at EBI (http://www.ebi.ac.uk). The results of this search are summarised in Table 3.4 and examples are given for IlvD, Pur and Tpi, in Figures 3.5 to 3.7.

Three non-conserved amino acid sequence changes were discovered. Allele IlvD had an amino acid change from M (methionine) in IlvD-8 and IlvD-9 to T (threonine) in IlvD-1, IlvD-7, IlvD-14 and IlvD-16 (Figure 3.5). Methionine contains a sulphur group and threonine has a hydroxyl group and is hydrophilic. This amino acid change resulted in a strand developing into a helix and was the only predicted conformational change observed in this dataset. An examination of all 88 STs showed that both the methionine and threonine amino acids occurred frequently, with methionine present in 31 of the STs while threonine occurred 57 times. The Pur allele had one amino acid change and this change was non-conserved (Figure 3.6), from R (arginine), present in most strains to C (cysteine), present only in Pur-5 (*B. thuringiensis* morrisoni). Arginine is basic, with two amino groups in a long chain, while cysteine has an –SH group. Allele Tpi had three amino acid changes (Figure 3.7), but only one was non-conserved, from E (glutamic acid) in Tpi-7 and Tpi-13 (found in several *B. thuringiensis* serovars) to A (alanine) in Tpi-1 (*B. anthracis*). Glutamic acid has a terminal carboxylate group while alanine is a small aliphatic amino acid with a methyl group.

There were four semi-conserved amino-acid changes, none of which produced a change in the predicted secondary structure of the protein. Three changes occurred in GlpF and two related only to *B. anthracis* strains. A (alanine) in all strains except *B. anthracis* changed to V (valine) in *B. anthracis* and S (serine) in all strains except *B. anthracis* changed to G (glycine) in *B. anthracis*. The third change in GlpF was from V (valine) to

A (alanine). Alanine, valine and glycine are aliphatic amino acids, while serine is also aliphatic but with a hydroxyl group. The only other semi-conserved amino acid change occurred in Gmk: - S (serine) to N (asparagine). Asparagine has an amide group. Additionally, eight conserved amino acid changes were found, two in Gmk, one in IlvD, four in PycA and one in Tpi. No amino acid changes were found in Pta.

Amino acid changes are rare events in the 'housekeeping' genes selected for MLST. Changes resulting in an altered protein conformation are extremely rare events, and only one such occurrence was detected, in the IlvD gene. This is a reflection of the fact that conservation of these genes is necessary for cell metabolism and non-synonymous mutations could result in non-viable progeny.

Annotated secondary structure diagram (PDBsum - SAS)

The predicted structure diagram shows a plot of the pasted amino acid sequence plus structural annotations from the structures included in an alignment from the proteins currently present in the database. The secondary structure shown in red has been derived from the closest homologue in the alignment, while the remaining parts (shown in varying shades of green) correspond to predicted secondary structure obtained using the DSC program (King *et al.*, 1997). The darker the colour, the more confident was the prediction.

Keys to the clustalW alignments and the secondary structure diagrams are given following Figure 3.7.

Table 3.4 Amino-acid changes in frequently occurring alleles

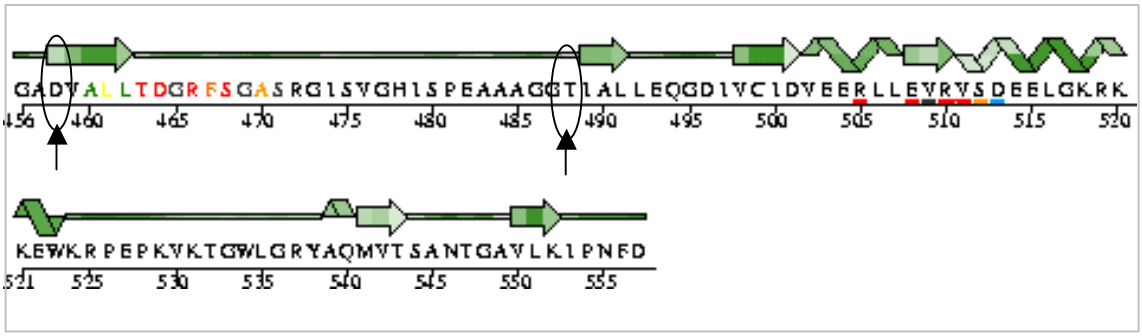
Allele	Number of alleles compared	*Alleles*	Number of amino-acid changes	Amino-acid change *(alleles)*	Type of change
GlpF	**5**	*1,3,8,12,15*	3	A (*3,8,12,15*) to V (*1*)	semi-conserved
				S (*3,8,12,15*) to G (*1*)	semi-conserved
				V (*12,15*) to A (*1,3,8*)	semi-conserved
Gmk	**5**	*1,2,7,8,9*	3	K (*1,2,8,9*) to R (*7*)	conserved
				S (*7,8,9*) to N (*1,2*)	semi-conserved
				D (*1,2*) to E (*7,8,9*)	conserved
IlvD	**6**	*1,7,8,9,14,16*	2	T (*1,7,14,16*) to M (*8,9*)	not conserved
				D (*1,7*) to E (*8*,9,*14,16*)	conserved
Pta	**6**	*1,2,5,11,12, 13*	none	none	-
Pur	**6**	*1,2,5,7,9,16*	1	R (*1,2,7,9,16*) to C (*5*)	not conserved
PycA	**5**	*1,3,8,12,16*	5	A (*3.8,12,16*) to V (*1*)	semi-conserved
				A (*8,12,16*) to S (*1,3*)	conserved
				M (*1,3,12,16*) to L (*8*)	conserved
				E (*8,12,16*) to K (*1,3*)	conserved
				K (*8,12,16*) to E (*1,3*)	conserved
Tpi	**3**	*1,7,13*	3	I (*1,7*) to L (*13*)	conserved
				V (*1*) *to* A *7,13*)	semi-conserved
				A (*1*) to E (*7,13*)	not conserved

IlvD

```
ilv14_          EEGGLRILKGNLAKDGAVIKSGATEVKRFEGPCVIFNSQDEALAGIMLGKVKKGDVVVIR 60
ilv16_          EEGGLRILKGNLAKDGAVIKSGATEVKRFEGPCVIFNSQDEALAGIMLGKVKKGDVVVIR 60
ilv8_           EEGGLRILKGNLAKDGAVIKSGATEVKRFEGPCVIFNSQDEALAGIMLGKVKKGDVVVIR 60
ilv9_           EEGGLRILKGNLAKDGAVIKSGATEVKRFEGPCVIFNSQDEALAGIMLGKVKKGDVVVIR 60
ilv1_           EEGGLRILKGNLAKDGAVIKSGATEVKRFEGPCVIFNSQDEALAGIMLGKVKKGDVVVIR 60
ilv7_           EEGGLRILKGNLAKDGAVIKSGATEVKRFEGPCVIFNSQDEALAGIMLGKVKKGDVVVIR 60
                ************************************************************

ilv14_          YEGPRGGPGMPEMLAPTSAIAGMGLGAEVALLTDGRFSGASRGISVGHISPEAAAGGTIA 120
ilv16_          YEGPRGGPGMPEMLAPTSAIAGMGLGAEVALLTDGRFSGASRGISVGHISPEAAAGGTIA 120
ilv8_           YEGPRGGPGMPEMLAPTSAIAGMGLGAEVALLTDGRFSGASRGISVGHISPEAAAGGMIA 120
ilv9_           YEGPRGGPGMPEMLAPTSAIAGMGLGAEVALLTDGRFSGASRGISVGHISPEAAAGGMIA 120
ilv1_           YEGPRGGPGMPEMLAPTSAIAGMGLGADVALLTDGRFSGASRGISVGHISPEAAAGGTIA 120
ilv7_           YEGPRGGPGMPEMLAPTSAIAGMGLGADVALLTDGRFSGASRGISVGHISPEAAAGGTIA 120
                **************************:*****************************  **

ilv14_          LLEQGDIVCID 131
ilv16_          LLEQGDIVCID 131
ilv8_           LLEQGDIVCID 131
ilv9_           LLEQGDIVCID 131
ilv1_           LLEQGDIVCID 131
ilv7_           LLEQGDIVCID 131
```



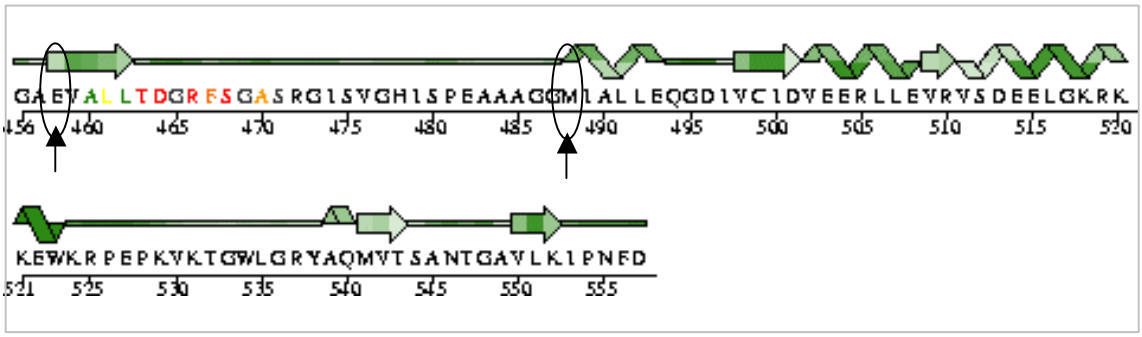


Figure 3.5 Alignment of amino acid sequence of part of the *ilvD* gene and predicted protein structure.

In the *ilvD* allele, there were two amino acid changes: -

conserved D (1,7) to E (8,9,14,16)

not conserved T (1,7,14,16) to M (8,9). This change resulted in a strand developing into a short helix.

Pur

```
pur1_           KRKLAAKVFRHTAAYDALISNYLTEQMGEESPETLTVTFEKKQDLRYGENPHQKATFYKA 60
pur2_           KRKLAAKVFRHTAAYDALISNYLTEQMGEESPETLTVTFEKKQDLRYGENPHQKATFYKA 60
pur9_           KRKLAAKVFRHTAAYDALISNYLTEQMGEESPETLTVTFEKKQDLRYGENPHQKATFYKA 60
pur16_          KRKLAAKVFRHTAAYDALISNYLTEQMGEESPETLTVTFEKKQDLRYGENPHQKATFYKA 60
pur7_           KRKLAAKVFRHTAAYDALISNYLTEQMGEESPETLTVTFEKKQDLRYGENPHQKATFYKA 60
pur5_           KCKLAAKVFRHTAAYDALISNYLTEQMGEESPETLTVTFEKKQDLRYGENPHQKATFYKA 60
                * ************************************************************

pur1_           PFAATSSVAYAEQLHGKELSYNNINDADAALSIVKEFTEPAVVAVKHMNPCGVGVG 116
pur2_           PFAATSSVAYAEQLHGKELSYNNINDADAALSIVKEFTEPAVVAVKHMNPCGVGVG 116
pur9_           PFAATSSVAYAEQLHGKELSYNNINDADAALSIVKEFTEPAVVAVKHMNPCGVGVG 116
pur16_          PFAATSSVAYAEQLHGKELSYNNINDADAALSIVKEFTEPAVVAVKHMNPCGVGVG 116
pur7_           PFAATSSVAYAEQLHGKELSYNNINDADAALSIVKEFTEPAVVAVKHMNPCGVGVG 116
pur5_           PFAATSSVAYAEQLHGKELSYNNINDADAALSIVKEFTEPAVVAVKHMNPCGVGVG 116
                ***********************************************************
```

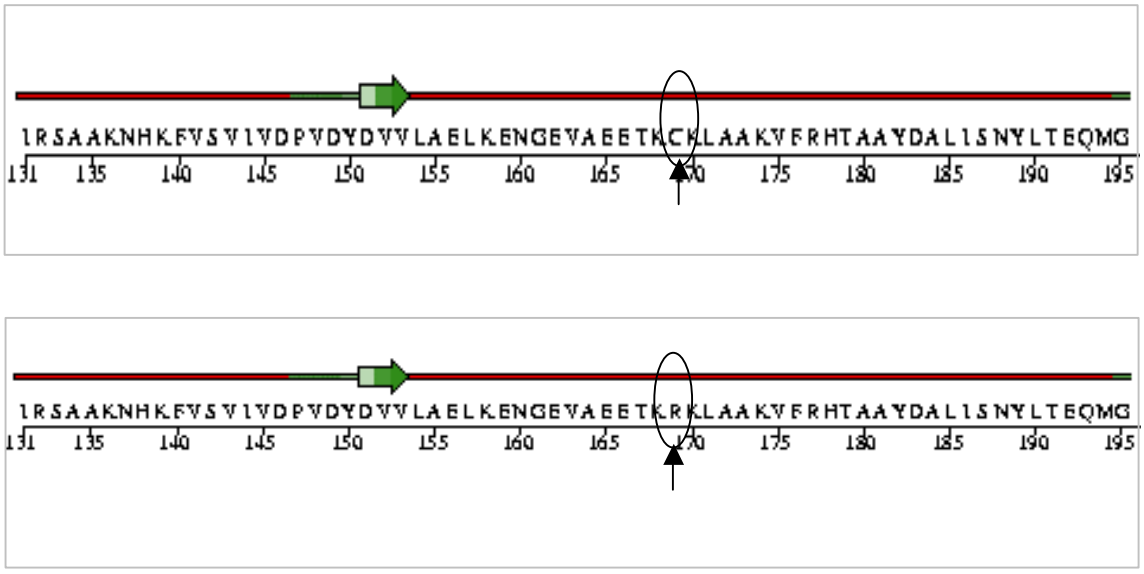


Figure 3.6 Alignment of amino acid sequences of part of the *pur* gene and predicted protein structure

In the *pur* allele, there was one amino acid change: -

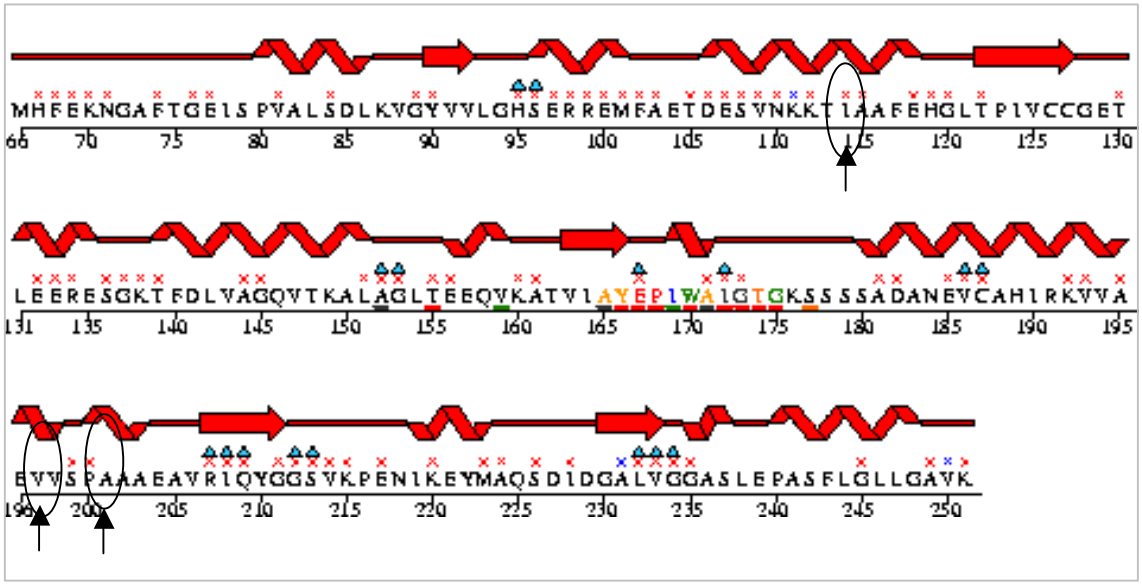
not conserved R (1, 2, 7,9, 16) to C (5). This change did not affect the predicted structure.

Tpi

```
tpi7_           ESVNKKTIAAFEHGLTPIVCCGETLEERESGKTFDLVAGQVTKALAGLTEEQVKATVIAY 60
tpi13_          ESVNKKTLAAFEHGLTPIVCCGETLEERESGKTFDLVAGQVTKALAGLTEEQVKATVIAY 60
tpi1_           ESVNKKTIAAFEHGLTPIVCCGETLEERESGKTFDLVAGQVTKALAGLTEEQVKATVIAY 60
                *******:****************************************************

tpi7_           EPIWAIGTGKSSSSADANEVCAHIRKVVAEAVSPEAAEAVRIQYGGSVKPENIKEYMAQS 120
tpi13_          EPIWAIGTGKSSSSADANEVCAHIRKVVAEAVSPEAAEAVRIQYGGSVKPENIKEYMAQS 120
tpi1_           EPIWAIGTGKSSSSADANEVCAHIRKVVAEVVSPAAAEAVRIQYGGSVKPENIKEYMAQS 120
                ****************************.*** ************************

tpi7_           DIDGALVGGASLEPASFLGLLGAVK 145
tpi13_          DIDGALVGGASLEPASFLGLLGAVK 145
tpi1_           DIDGALVGGASLEPASFLGLLGAVK 145
                *************************
```



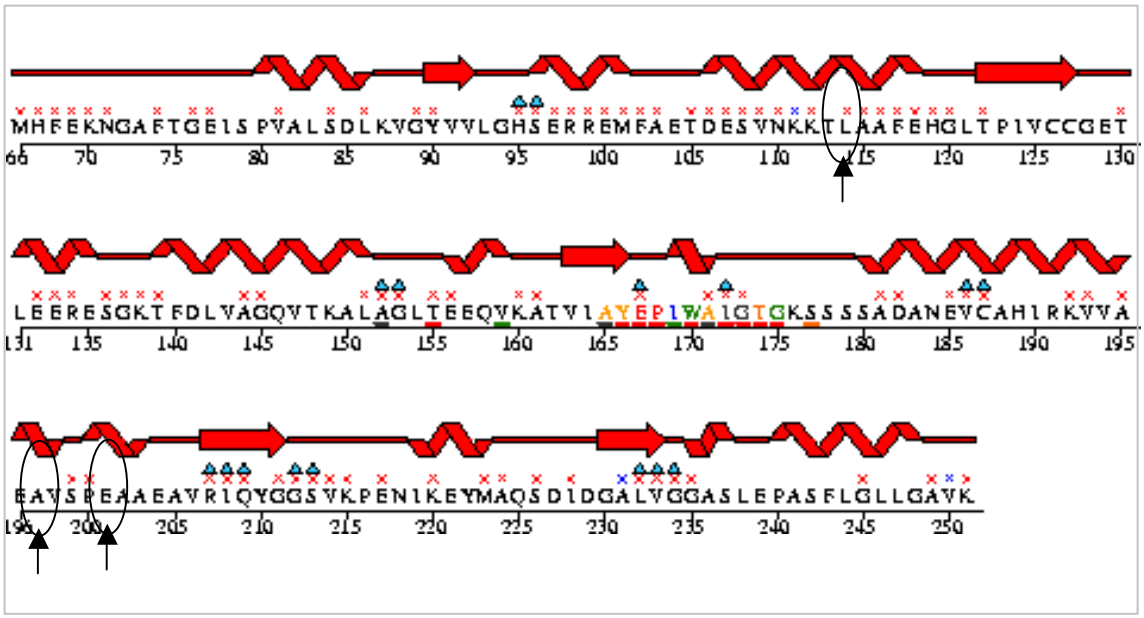


Figure 3.7 Alignment of amino acid sequence of part of the *tpi* gene and predicted protein structure

In the *tpi* allele, there were three amino acid changes: -

Conserved: - I (1,7) to L (13)

semi conserved: - V (1) to A (7,13)

not conserved: - A (1) to E (7,13)

None of these changes affected the predicted secondary structure of the protein.

Alignment key ClustalW (1.8)

This key is required to interpret the colour and symbols shown in the amino acid alignments (Figures 3.5 to 3.7)

The colouring of residues follows the physiochemical criteria:

AVFPMILW RED Small (small+ hydrophobic (incl.aromatic -Y))
DE BLUE Acidic
RHK MAGENTA Basic
STYHCNGQ GREEN Hydroxyl + Amine + Basic - Q

CONSENSUS SYMBOLS:

An alignment will display by default the following symbols denoting the degree of conservation observed in each column:

"*" means that the residues or nucleotides in that column are identical in all sequences in the alignment.

":" means that conserved substitutions have been observed, according to the COLOUR table above.

"." means that semi-conserved substitutions are observed.

The data above are reproduced from the ClustalW help page.

Secondary structure key (PDBsum – SAS)

Secondary structure assigned from closest homologue: -

Helix  Strand  Coil

3.1.3 Phylogenetic structure of the *B. cereus* group

3.1.3.i Basic division into clades

The nucleotide sequence data for all seven alleles were concatenated producing a sequence length of 2838 bases for each ST. The sequences were aligned and a neighbour joining (NJ) tree constructed using ClustalW. The resulting dendrogram file was visualised using Treeview, version 1.6.6 (r.page@bio.gla.ac.uk). The NJ tree was saved as an 'emf' file and inserted into Microsoft Word for annotation. The resultant unrooted NJ tree is shown in Figure 3.8 and a tree rooted with *B. pseudomycoides* as the outgroup is shown in Figure 3.9.

The unrooted tree was divided into four groups or clades. Clade 1 comprised *B. anthracis*, the majority of the *B. cereus* strains, and some unusual *B. thuringiensis* strains. Clade 2 was made up of *B. thuringiensis* and *B. cereus* strains. Clade 3 contained *B. mycoides, B. weihenstephanensis*, and some *B. cereus* strains which were isolated from Norwegian soil. Finally, the 'outliers' group comprised only two strains, a *B. pseudomycoides* and a strain which was received as *B. mycoides*. Since the outliers consisted of only two STs it was not allocated clade status in this study. However, as these STs were so far removed from the other clades, one of them, ST-83 *B. pseudomcoides*, was selected as the 'root' for the phylogenetic tree. STs within the clades were divided into groups, and these were defined as lineages in Figure 3.9.

3.1.3.ii Bootstrap values

In the main paper derived from this thesis, 'Population Structure and Evolution of the *Bacillus cereus* Group' (Priest *et al.*, 2004), a 'maximum likelihood' phylogenetic tree was included for 59 STs. The branches leading to Clades 1 and 2 were strongly defined, with bootstrap values greater than 85% (a highly divergent strain, ST-9 was excluded). For the 88 STs studied here, the division was not so well supported. This was probably due to the inclusion in the dataset of more atypical strains, such as those assigned to ST-72, ST-74, ST-78, ST-80 and ST-82, the allocation of which to clades was poorly supported by the bootstrap analysis. However, the primary structure was still retained.

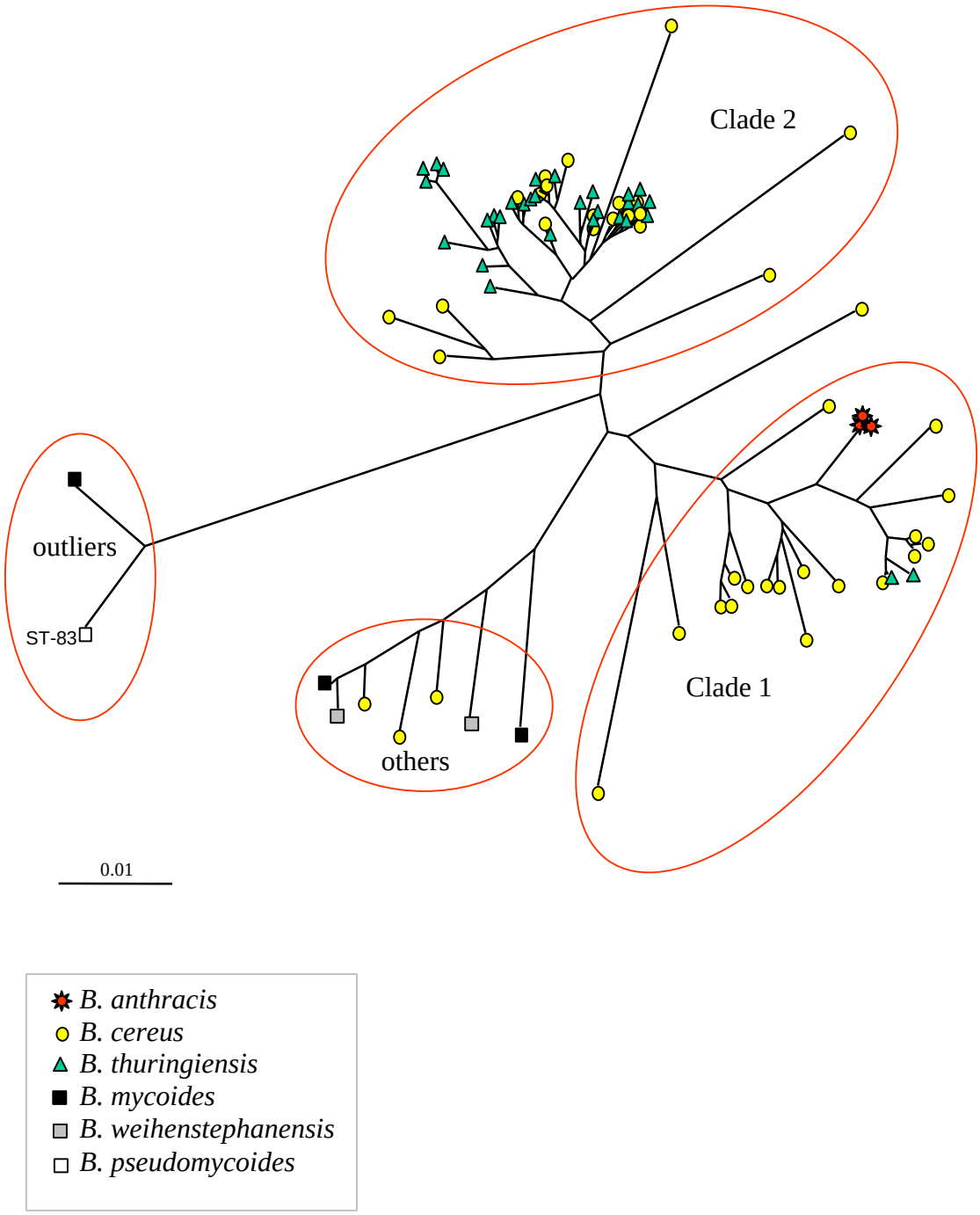


Figure 3.8 Unrooted NJ tree constructed from the concatenated sequences of seven genes from 88 Sequence Types of the *Bacillus cereus* group. (The concatenated sequence is 2838 bp).

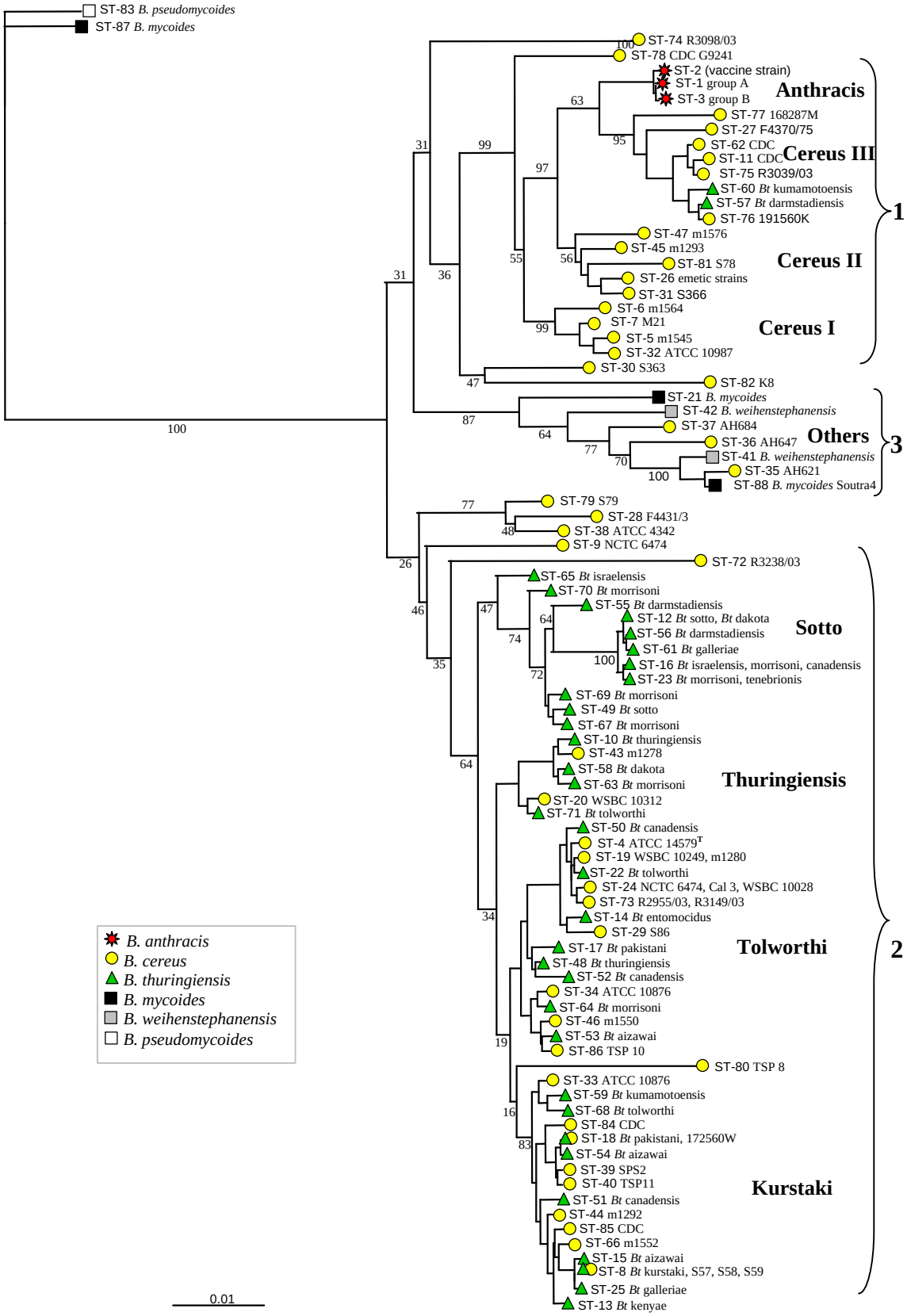


Figure 3.9 Rooted NJ tree constructed from the concatenated sequences of 88 STs (146 strains) of the *B. cereus* group (rooted with ST-83, *B. pseudomycoides*). Bootstrap values are given as a percentage (1000 trials).

3.1.4 Phylogenetic structure of the lineages

Figure 3.9 shows that the STs could be divided into three clades, with a total of nine lineages. Clade 1 was composed of lineages Anthracis, Cereus I, which included ATCC 10987, an isolate from milk for which the full genome sequence is available (Rasko *et al.,* 2004). Cereus II contained the emetic-toxin producing strains. Cereus III, was the closest lineage to Anthracis and included two rare *B. thuringiensis* strains. Clade 2 was comprised mostly of *B. thuringiensis* strains with *B. cereus* intermixed. The type strains of *B. cereus* and *B. thuringiensis* were recovered in clade 2. The lineages of clade 2 were labelled according to serovars Thuringiensis, Sotto, Tolworthi and Kurstaki. Clade 3 was made up of *B. mycoides, B. weihenstephanensis* and three Norwegian soil isolates but no intra-clade lineages were apparent. There was strong (>80%) bootstrap support for the following lineages: - Anthracis, Cereus I, Cereus III, Kurstaki, sub-group within Sotto and the 'Others'.

Evaluation of species, clades and lineages

B. anthracis, B. cereus and *B. thuringiensis* are currently defined as separate species and the STs were divided in accordance with this designation. The proposed divisions into four clades and nine lineages were also examined. The validity of species, clade and lineage designations was investigated by calculating the values for Fst, fixed differences and shared mutations for each category of division. The values were calculated using the DnaSP program (Hudson *et al.*, 1992)**.**

Fst measures gene flow between two populations, If there was no difference between two populations, the value of Fst would be 0 and if the two populations were very different, the value of Fst would be close to 1.

Fixed Differences are nucleotide sites at which all of the sequences in one population are different from all of the sequences in the second population.

Shared Mutations are the total number of mutations that are shared between the two populations.

3.1.4.i STs separated by 'species' (as currently defined)

B. anthracis, B. cereus and *B. thuringiensis* are currently defined as separate species, and the data were divided in accordance with this designation: - *B. anthracis* (3 STs), *B. cereus* (44 STs), *B. thuringiensis* (34 STs), (STs 8 and 18 are found in both *B. cereus* and *B. thuringiensis*). Strains that did not fit into the three 'species' formed two groups – the others (seven STs comprising two *B. mycoides*, two *B. weihenstephanensis* and three strains from Norway), and the 'outliers' (two STs - one *B. pseudomycoides* and one other strain, received as *B. mycoides*).

Fst

Comparing *B. cereus* with *B. thuringiensis* STs, the vaue of Fst for the concatenated data was 0.162, The lowest value was for the *glpF* gene, at 0.085, and the highest was for *gmk*, at 0.262 , as shown in Table 3.5 (Fst values < 0.25 are highlighted). This indicates that there is very little difference between the two populations. Fst for *B. anthracis* against *B. cereus* was 0.524 for the concatenated data, and between 0.463 and 0.587 for the individual alleles. When *B. anthracis* was compared with *B. thuringiensis*, the Fst value was much higher, at 0.783 for the concatenated data, and ranged from 0.662 to 0.891 for the alleles. This implies that there is less gene flow (more differences) between strains of *B. anthracis* and *B. thuringiensis* than there is between strains of *B. anthracis* and *B. cereus*. An interesting Fst result arose when *B. cereus* was compared with the outliers (*B. pseudomycoides*), and a low Fst value of 0.135 for *glpF* was found, implying that there is little difference between these two populations for the *glpF* allele. Similarly, a very low Fst value for the *ilvD* allele (0.038) was noticed between *B. thuringiensis* and the outliers.

The low value for Fst confirms that there is no justification for considering *Bacillus cereus* and *Bacillus thuringiensis* as separate species, other than the observation that *B. thuringiensis* strains produce a crystal protein at the time of sporulation. However, *B. anthracis* could be considered to be a separate species from *B. cereus, B. thuringiensis*, the 'others' and the 'outliers'.

Fixed differences

There were no fixed differences between *B. cereus* and *B. thuringiensis strains* using the concatenated data shown in Table 3.6 (fixed differences of 0 are highlighted). Five fixed differences were found between *B. anthracis* and *B. cereus* strains and eight fixed differences between *B. anthracis* and *B. thuringiensis* strains indicating that *B. anthracis* is evolving as a separate lineage.

Shared mutations

There were 246 shared mutations between *B. cereus* and *B. thuringiensis* strains on the basis of the concatenated data. There was one shared mutation between *B. anthracis* and *B. cereus*, and one shared mutation between *B. anthracis* and *B. thuringiensis* strains. This mutation occurred in the *pycA* allele (Table 3.7).

In summary, the low values for 'Fst', no 'fixed differences', combined with the presence of many 'shared mutations' emphasises that there is no justification in describing *B. cereus* and *B. thuringiensis* as separate species.

Table 3.5 Fst for data separated by species.

Fst concatenated data

	B .thuringiensis	others	outliers	*B. anthracis*
B. cereus	**0.162**	**0.508**	**0.665**	**0.524**
B. thuringiensis		**0.634**	**0.741**	**0.783**
others			**0.762**	**0.818**
outliers				**0.898**

Fst *glpF*

	B .thuringiensis	others	outliers	*B. anthracis*
B. cereus	**0.085**	**0.764**	**0.135**	**0.556**
B. thuringiensis		**0.846**	**0.425**	**0.760**
others			**0.857**	**0.925**
outliers				**0.522**

Fst *gmk*

	B .thuringiensis	others	outliers	*B. anthracis*
B. cereus	**0.262**	**0.554**	**0.850**	**0.517**
B. thuringiensis		**0.717**	**0.935**	**0.891**
others			**0.889**	**0.833**
outliers				**0.981**

Fst *ilvD*

	B .thuringiensis	others	outliers	*B. anthracis*
B. cereus	**0.116**	**0.155**	**0.270**	**0.509**
B. thuringiensis		**0.313**	**0.038**	**0.768**
others			**0.460**	**0.612**
outliers				**0.908**

Fst *pta*

	B .thuringiensis	others	outliers	*B. anthracis*
B. cereus	**0.090**	**0.585**	**0.545**	**0.587**
B. thuringiensis		**0.637**	**0.544**	**0.690**
others			**0.621**	**0.814**
outliers				**0.737**

Fst *pur*

	B .thuringiensis	others	outliers	*B. anthracis*
B. cereus	**0.095**	**0.495**	**0.735**	**0.463**
B. thuringiensis		**0.656**	**0.820**	**0.662**
others			**0.779**	**0.804**
outliers				**0.918**

Fst *pycA*

	B .thuringiensis.	others	outliers	*B. anthracis*
B. cereus	**0.250**	**0.558**	**0.754**	**0.533**
B. thuringiensis		**0.692**	**0.836**	**0.821**
others			**0.886**	**0.887**
outliers				**0.973**

Fst *tpi*

	B .thuringiensis	others	outliers	*B. anthracis*
B. cereus	**0.118**	**0.680**	**0.554**	**0.581**
B. thuringiensis		**0.663**	**0.566**	**0.710**
others			**0.632**	**0.948**
outliers				**0.688**

Table 3.6 Fixed differences for data separated by species

Fixed differences concatenated data

	B. thuringiensis	others	outliers	*B. anthracis*
B. cereus	**0**	**12**	**72**	**5**
B. thuringiensis		**26**	**97**	**8**
others			**132**	**100**
outliers				**214**

Fixed differences *glpF*

	B. thuringiensis	others	outliers	*B. anthracis*
B. cereus	**0**	**1**	**0**	**2**
B. thuringiensis		**12**	**0**	**2**
others			**15**	**19**
outliers				**2**

Fixed differences *gmk*

	B. thuringiensis	others	outliers	*B. anthracis*
B. cereus	**0**	**0**	**29**	**0**
B. thuringiensis		**0**	**39**	**0**
others			**37**	**13**
outliers				**53**

Fixed differences *ilvD*

	B. thuringiensis	others	outliers	*B. anthracis*
B. cereus	**0**	**0**	**0**	**1**
B. thuringiensis		**0**	**0**	**2**
others			**0**	**8**
outliers				**38**

Fixed differences *pta*

	B. thuringiensis	others	outliers	*B. anthracis*
B. cereus	**0**	**4**	**6**	**0**
B. thuringiensis		**3**	**9**	**0**
others			**15**	**8**
outliers				**21**

Fixed differences *pur*

	B. thuringiensis	others	outliers	*B. anthracis*
B. cereus	**0**	**1**	**15**	**0**
B. thuringiensis		**4**	**21**	**0**
others			**26**	**18**
outliers				**40**

Fixed differences *pycA*

	B. thuringiensis	others	outliers	*B. anthracis*
B. cereus	**0**	**4**	**17**	**1**
B. thuringiensis		**5**	**23**	**2**
others			**30**	**24**
outliers				**49**

Fixed differences *tpi*

	B. thuringiensis	others	outliers	*B. anthracis*
B. cereus	**0**	**2**	**5**	**1**
B. thuringiensis		**2**	**5**	**2**
others			**9**	**10**
outliers				**11**

Table 3.7 Shared mutations for data separated by species

Shared mutations concatenated data

	B. thuringiensis	others	outliers	*B. anthracis*
B. cereus	**246**	**117**	**24**	**1**
B. thuringiensis		**76**	**20**	**1**
others			**9**	**1**
outliers				**0**

Shared mutations *glpF*

	B. thuringiensis	others	outliers	*B. anthracis*
B. cereus	**16**	**5**	**3**	**0**
B. thuringiensis		**1**	**2**	**0**
others			**0**	**0**
outliers				**0**

Shared mutations *gmk*

	B. thuringiensis	others	outliers	*B. anthracis*
B. cereus	**29**	**31**	**0**	**0**
B. thuringiensis		**13**	**0**	**0**
others			**0**	**0**
outliers				**0**

Shared mutations *ilvD*

	B. thuringiensis	others	outliers	*B. anthracis*
B. cereus	**65**	**48**	**7**	**0**
B. thuringiensis		**40**	**7**	**0**
others			**5**	**0**
outliers				**0**

Shared mutations *pta*

	B. thuringiensis	others	outliers	*B. anthracis*
B. cereus	**26**	**4**	**6**	**0**
B. thuringiensis		**2**	**5**	**0**
others			**2**	**0**
outliers				**0**

Shared mutations *pur*

	B. thuringiensis	others	outliers	*B. anthracis*
B. cereus	**39**	**14**	**5**	**0**
B. thuringiensis		**9**	**3**	**0**
others			**2**	**0**
outliers				**0**

Shared mutations *pycA*

	B. thuringiensis	others	outliers	*B. anthracis*
B. cereus	**52**	**14**	**0**	**1**
B. thuringiensis		**10**	**0**	**1**
others			**0**	**1**
outliers				**0**

Shared mutations *tpi*

	B. thuringiensis	others	outliers	*B. anthracis*
B. cereus	**19**	**1**	**3**	**0**
B. thuringiensis		**1**	**3**	**0**
others			**0**	**0**
outliers				**0**

3.1.4.ii STs separated by clades

Fst

Comparing clade 1 with clade 2, the value of Fst for the concatenated sequences was 0.561 (Table 3.8), indicating that there is considerable difference between the two clades. However, four of the seven individual alleles (*glpF, pta*, *pur* and *tpi)* had a Fst value of less than 0.4. The *glpF* allele was unusual, in the low Fst value of 0.258 for clade 1 against clade 2. A value of 0 was returned for clade 1 against the outliers, indicating that there was no difference between clade 1 and the outliers at this locus. Similarly for the *ilvD* locus, clade 2 against the outliers gave a Fst value of 0.012. This indicates horizontal exchange between the members of these clades at the specified loci.

Fixed differences

There were no fixed differences between clade 1 and clade 2 at any of the loci (Table 3.9). For the concatenated data, there were 102 fixed differences between clade 1 and the outliers, 97 fixed differences between clade 2 and the outliers, and 132 fixed differences between clade 3 and the outliers. These were not distributed across all alleles, and *glpF* had no fixed differences between clade 1 and the outliers, or between clade 2 and the outliers. Also, there were no fixed differences between clade 1 and clade 3, or between clade 2 and clade 3 for the *gmk* and *ilvD* loci. The absence of fixed differences between populations for particular alleles suggests that there has been some horizontal gene exchanges involving the *glpF, gmk* and *ilvD* alleles.

Shared mutations

There were 157 shared mutations between STs of clade 1 and clade 2. Shared mutations occurred in most of the clades and for most of the loci. The exceptions being *gmk* and *pycA,* where there were no shared mutations between the outliers and any of the three clades (Table 3.10).

In summary, some of the Fst values for individual genes were low, which may be evidence for recombination at these loci. Shared mutations were prevalent between most clades and there were no fixed differences between clade 1 and clade 2, indicating that while these clades may be separated, the distinction is not very strong.

Table 3.8 Fst for data separated by clades**.**

Fst concatenated data

	Clade 2	Clade 3	outliers
Clade 1	**0.561**	**0.596**	**0.751**
Clade 2		**0.653**	**0.754**
Clade 3			**0.762**

Fst *glpF*

	Clade 2	Clade 3	outliers
Clade 1	**0.258**	**0.755**	**0**
Clade 2		**0.835**	**0.411**
Clade 3			**0.857**

Fst *gmk*

	Clade 2	Clade 3	outliers
Clade 1	**0.722**	**0.697**	**0.917**
Clade 2		**0.715**	**0.934**
Clade 3			**0.889**

Fst *ilvD*

	Clade 2	Clade 3	outliers
Clade 1	**0.619**	**0.367**	**0.721**
Clade 2		**0.364**	**0.012**
Clade 3			**0.460**

Fst *pta*

	Clade 2	Clade 3	outliers
Clade 1	**0.375**	**0.580**	**0.600**
Clade 2		**0.663**	**0.555**
Clade 3			**0.621**

Fst *pur*

	Clade 2	Clade 3	outliers
Clade 1	**0.341**	**0.409**	**0.691**
Clade 2		**0.685**	**0.840**
Clade 3			**0.779**

Fst *pycA*

	Clade 2	Clade 3	outliers
Clade 1	**0.689**	**0.750**	**0.886**
Clade 2		**0.703**	**0.845**
Clade 3			**0.886**

Fst *tpi*

	Clade 2	Clade 3	outliers
Clade 1	**0.285**	**0.668**	**0.510**
Clade 2		**0.707**	**0.590**
Clade 3			**0.632**

Table 3.9 Fixed Differences for data separated by clades

Fixed differences concatenated data

	Clade 2	Clade 3	outliers
Clade 1	**0**	**14**	**102**
Clade 2		**31**	**97**
Clade 3			**132**

Fixed differences *glpF*

	Clade 2	Clade 3	outliers
Clade 1	**0**	**1**	**0**
Clade 2		**10**	**0**
Clade 3			**15**

Fixed differences *gmk*

	Clade 2	Clade 3	outliers
Clade 1	**0**	**0**	**34**
Clade 2		**0**	**38**
Clade 3			**37**

Fixed differences *ilvD*

	Clade 2	Clade 3	outliers
Clade 1	**0**	**0**	**12**
Clade 2		**0**	**0**
Clade 3			**0**

Fixed differences *pta*

	Clade 2	Clade 3	outliers
Clade 1	**0**	**3**	**9**
Clade 2		**5**	**9**
Clade 3			**15**

Fixed differences *pur*

	Clade 2	Clade 3	outliers
Clade 1	**0**	**1**	**16**
Clade 2		**7**	**21**
Clade 3			**26**

Fixed differences *pycA*

	Clade 2	Clade 3	outliers
Clade 1	**0**	**7**	**26**
Clade 2		**7**	**23**
Clade 3			**30**

Fixed differences *tpi*

	Clade 2	Clade 3	outliers
Clade 1	**0**	**2**	**5**
Clade 2		**2**	**6**
Clade 3			**9**

Table 3.10 Shared mutations for data separated by clades

Shared mutations concatenated data

	Clade 2	Clade 3	outliers
Clade 1	**157**	**79**	**16**
Clade 2		**81**	**20**
Clade 3			**9**

Shared mutations *glpF*

	Clade 2	Clade 3	outliers
Clade 1	**13**	**4**	**3**
Clade 2		**1**	**2**
Clade 3			**0**

Shared mutations *gmk*

	Clade 2	Clade 3	outliers
Clade 1	**20**	**23**	**0**
Clade 2		**13**	**0**
Clade 3			**0**

Shared mutations *ilvD*

	Clade 2	Clade 3	outliers
Clade 1	**42**	**32**	**4**
Clade 2		**42**	**7**
Clade 3			**5**

Shared mutations *pta*

	Clade 2	Clade 3	outliers
Clade 1	**15**	**3**	**4**
Clade 2		**3**	**6**
Clade 3			**2**

Shared mutations *pur*

	Clade 2	Clade 3	outliers
Clade 1	**31**	**11**	**3**
Clade 2		**9**	**3**
Clade 3			**2**

Shared mutations *pycA*

	Clade 2	Clade 3	outliers
Clade 1	**24**	**5**	**0**
Clade 2		**12**	**0**
Clade 3			**0**

Shared mutations *tpi*

	Clade 2	Clade 3	outliers
Clade 1	**12**	**1**	**2**
Clade 2		**1**	**2**
Clade 3			**0**

3.1.4.iii STs separated by lineages

Nine lineages are recognised on the NJ tree shown in Figure 3.9. The lineages were named Anthracis, Cereus I, Cereus II, Cereus III, Kurstaki, Tolworthi, Sotto, Thuringiensis and Others (the 'Outliers' group consisting of only two strains, was not considered a lineage, but is included here for completeness).

Fst

Most of the Fst values for the data separated by lineages were above 0.5, indicating that the populations are different from each other. There were two borderline exceptions; Cereus I compared with Cereus II, where the Fst value was 0.476, and Kurstaki against Tolworthi, where the Fst value was 0.475 (Table 3.11).

Fixed Differences

There were fixed differences between all lineages, supporting their distinction (Table 3.12).

Shared Mutations

There was only one shared mutation between Anthracis and another lineage, the 'others' (Table 3.13). There were shared mutations between the other lineages, with the exception of Thuringiensis and the 'outliers'. The highest number of shared mutations was between Sotto and the 'others', where 28 shared mutations were recorded, and 22 between Sotto and Tolworthi.

Summary

The consistently high values for Fst between lineages, together with large numbers of fixed differences and few shared mutations confirmed that dividing the data into lineages, as suggested by Figure 3.9, was appropriate for this dataset. Details of the strains and their lineage allocation are given in Table 3.14.

The statistical tests carried out confirmed that classification of the *B. cereus* group of bacteria by the currently defined 'species' designation is inappropriate and that division according to three major clades, sub-divided into nine (or ten) lineages is more suitable.

Table 3.11 Fst for lineages across all loci using 78 Sequence Types
*****(10 STs represented diverse strains and were not assigned to lineages)

No STs	Lineage	Cereus I	Cereus II	Cereus III	Kurstaki	Tolworthi	Sotto	Thuring.	Others	Outliers
3	Anthracis	0.849	0.681	0.690	0.939	0.923	0.901	0.936	0.818	0.898
4	Cereus I		0.476	0.740	0.860	0.838	0.825	0.853	0.757	0.861
5	Cereus II			0.570	0.826	0.810	0.798	0.820	0.731	0.833
8	Cereus III				0.874	0.861	0.844	0.872	0.741	0.848
16	Kurstaki					0.475	0.686	0.611	0.766	0.840
16	Tolworthi						0.644	0.528	0.759	0.836
11	Sotto							0.602	0.736	0.824
6	Thuringiensis								0.764	0.838
7	Others									0.762
2	Outliers									

*Unassigned STs ST-9, ST-28, ST-30, ST-38, ST-72, ST-74, ST-78, ST-79, ST-80 and ST-82

Table 3.12 Fixed differences for lineages across all loci using 78 Sequence Types (10 STs are undefined)

No STs	Lineage	Cereus I	Cereus II	Cereus III	Kurstaki	Tolworthi	Sotto	Thuring.	Others	Outliers
3	Anthracis	54	28	24	125	120	114	123	100	214
4	Cereus I		11	44	93	88	83	91	78	199
5	Cereus II			21	93	91	90	91	79	184
8	Cereus III				117	114	102	118	72	184
16	Kurstaki					3	20	8	63	156
16	Tolworthi						19	7	67	160
11	Sotto							18	63	156
6	Thuringiensis								69	156
7	Others									132
2	Outliers									

Table 3.13 Shared mutations for lineages across all loci using 78 Sequence Types (10 STs are undefined)

No STs	Lineage	Cereus I	Cereus II	Cereus III	Kurstaki	Tolworthi	Sotto	Thuring.	Others	Outliers
3	Anthracis	0	0	0	0	0	0	0	1	0
4	Cereus I		8	7	4	3	2	4	5	3
5	Cereus II			18	8	5	9	2	13	6
8	Cereus III				6	8	11	4	17	5
16	Kurstaki					6	17	1	11	9
16	Tolworthi						22	7	10	3
11	Sotto							18	28	6
6	Thuringiensis								8	0
7	Others									9
2	Outliers									

3.1.5 SplitsTrees of lineages

The division of STs into lineages was further investigated by the method of split decomposition (Huson, 1997). Each lineage was investigated separately, and the SplitsTree graphs produced for each lineage are shown in Figures 3.10 to 3.18. Tables are included for each SplitsTree graph showing the numbers of nucleotide changes between the STs and the alleles involved.

Lineage Anthracis (Figure 3.10)
There was only one nucleotide difference between ST-1 and ST-3, and two differences between ST-1 and ST-2, and this variation is reflected in the branch lengths shown.

Cereus I (Figure 3.11)
The close relationship of ST-5, ST-32 (*B. cereus* ATCC 10987) and ST-7 was apparent in the SplitsTree graph (Figure 3.11) All are *B. cereus* strain isolated from milk products. ST-6 (a strain of *B. cereus* isolated from dried coconut in Brazil) was more distant with polymorphisms in every allele.

Cereus II (Figure 3.12)
ST-26 (the emetic strains) were closely related to ST-31, representing a strain isolated from the North Sea. Other STs in this lineage included strains isolated from food such as chicken, cheese and rice (Table 3.14 and Appendix A)

Cereus III (Figure 3.13)
ST-11, ST-62 and ST-75 were very closely related, and formed a small sub-group. ST-27 and ST-77 appeared to be further removed. All are *B. cereus* strains:- ST-11 and ST-62 were submitted to the MLST database from CDC and probably represent clinical isolates of *B. cereus*. ST-75 and ST-77 were isolated from blood and a wound infection (Barker *et al.,* 2005). ST-27 (F4370/75) was isolated from barbecued chicken. There were 24 nucleotide changes between ST-57 (*Bt* darmstadiensis, T10024 from Pakistan) and ST-27 for the *ilvD* allele, suggesting that recombination may have occurred.

.

Sotto, **Thuringiensis and Tolworthi** (Figures 3.14, 3.15, 3.16)

Sotto, Thuringiensis and Tolworthi were each split into two sub-lineages. There were 26 to 30 nucleotide changes at the *ilvD* locus for the Sotto lineage indicating recombination. Similarly in the Tolworthi lineage, there were 14 to 16 nucleotide changes at the *ilvD* locus, again suggesting recombination at this locus. Lineage Thuringiensis had only 3 to 6 changes at the *ilvD* locus, but two STs which were distinct from the main group had 11 and 14 nucleotide changes at the *pycA* allele.

Kurstaki (Figure 3.17)

Kurstaki could be divided into three sub-lineages in the SplitsTree graph. The numbers of nucleotide changes between STs was fewer than 10 for all alleles, suggesting that mutation rather than recombination is responsible for evolutionary change in this lineage.

Others (Figure 3.18)

The others were composed of *B. mycoides*, *B. weihenstephanensis* and *B. cereus* strains from Norway. There were 30 or 31 nucleotide changes from ST-21 (*B. mycoides*) to all the other STs for the *gmk* allele, which strongly suggests recombination at this locus. Two *B. cereus* strains isolated from Norwegian soil, (ST-36 and ST-37) and the type strain of *B. weihenstephanensis* (ST-42) had 27 to 35 nucleotide changes at the *ilvD* locus and there may have been additional recombination at this allele. ST-88, a *B. mycoides* from Soutra in south Scotland, was very close to ST-35, a *B. cereus* from Norway, with few mutational changes as indicated.

Number of nucleotide changes in the Anthracis lineage							
ST change	***glpF***	***gmk***	***ilvD***	***pta***	***pur***	***pycA***	***tpi***
ST 1 to ST 2			1			1	
ST 1 to ST 3	1						

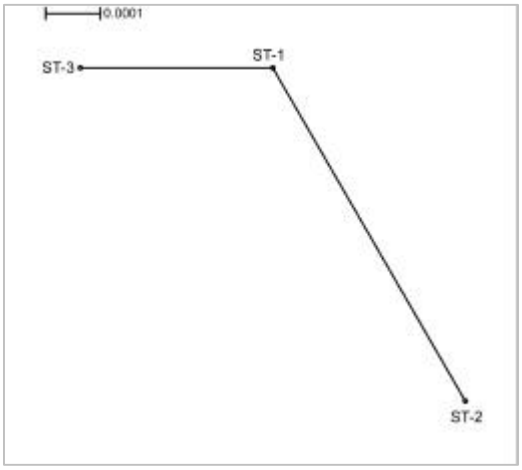


Figure 3.10 SplitsTree graph for the Anthracis lineage of clade 1 (bar = number of nucleotide changes per site)

Number of nucleotide changes in the Cereus I lineage							
ST change	***glpF***	***gmk***	***ilvD***	***pta***	***pur***	***pycA***	***tpi***
ST 32 to ST 5	3				2	3	1
ST 32 to ST 6	1	3	8	3	3	1	10
ST 32 to ST 7	7						

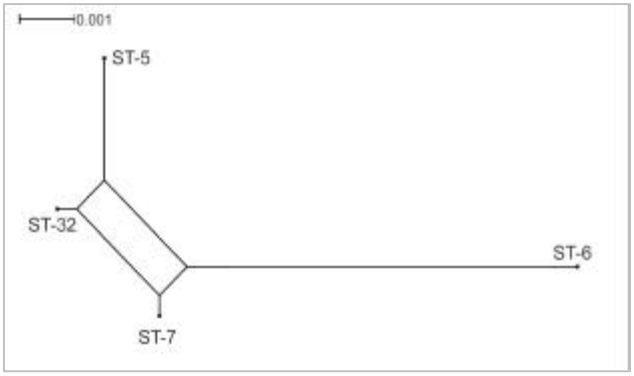


Figure 3.11 SplitsTree graph for the Cereus I lineage of clade 1 (bar = number of nucleotide changes per site)

Number of nucleotide changes in the Cereus II lineage							
ST change	***glpF***	***gmk***	***ilvD***	***pta***	***pur***	***pycA***	***tpi***
ST 26 to ST 31	4		6			1	2
ST 26 to ST 45	2	1	5		20		2
ST 26 to ST 47	5		17		13		3
ST 26 to ST 81			6		27	1	2

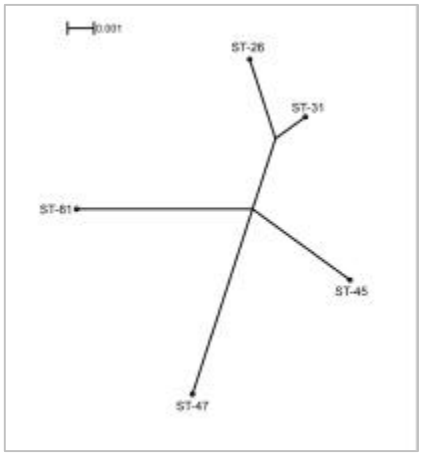


Figure 3.12 SplitsTree graph for the Cereus II lineage of clade 1 (bar = number of nucleotide changes per site)

Number of nucleotide changes in the Cereus III lineage							
ST change	***glpF***	***gmk***	***ilvD***	***pta***	***pur***	***pycA***	***tpi***
ST 57 to ST 11		1	4	2	2	1	
ST 57 to ST 27		1	24	6	1	2	
ST 57 to ST 60	2			1			8
ST 57 to ST 62	3	1	4	2	1	2	
ST 57 to ST 75	2	1	4	2	1	2	
ST 57 to ST 76				1			1
ST 57 to ST 77	4	2	15	9	5	2	7

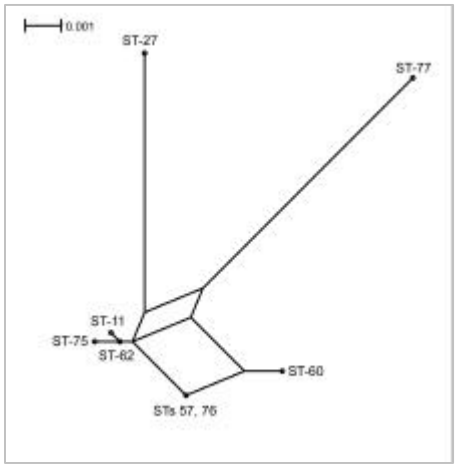


Figure 3.13 SplitsTree graph for the Cereus III lineage of clade 1 (bar = number of nucleotide changes per site)

Number of nucleotide changes in the Sotto lineage							
ST change	***glpF***	***gmk***	***ilvD***	***pta***	***pur***	***pycA***	***tpi***
ST 12 to ST 16					1	4	
ST 12 to ST 23					1	4	
ST 12 to ST 49			28				
ST 12 to ST 55			30			1	
ST 12 to ST 56						1	
ST 12 to ST 61				1		1	
ST 12 to ST 65			26		1	17	
ST 12 to ST 67			29		1		
ST 12 to ST 69			26		1	4	
ST 12 to ST 70			26		9	4	

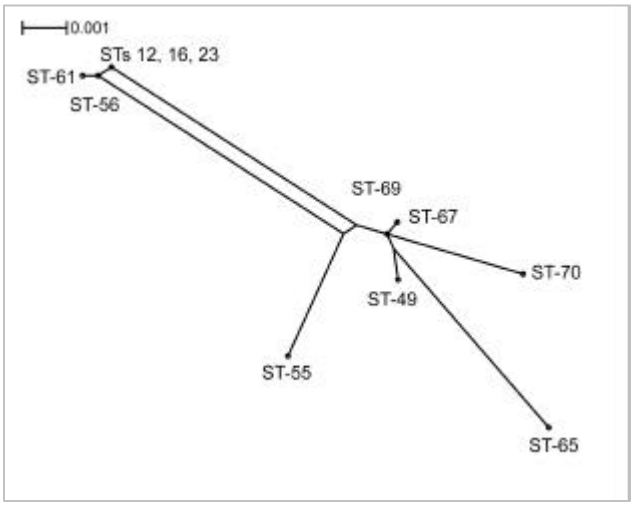


Figure 3.14 SplitsTree graph for the Sotto lineage of clade 2 (bar = number of nucleotide changes per site)

Number of nucleotide changes in the Thuringiensis lineage							
ST change	***glpF***	***gmk***	***ilvD***	***pta***	***pur***	***pycA***	***tpi***
ST 10 to ST 20	2		5	1	1	11	2
ST 10 to ST 43	1	1	3		1	2	1
ST 10 to ST 58	1		4	1	1		1
ST 10 to ST 63		1	4	2	1	1	3
ST 10 to ST 71	2	1	5	1	1	14	2

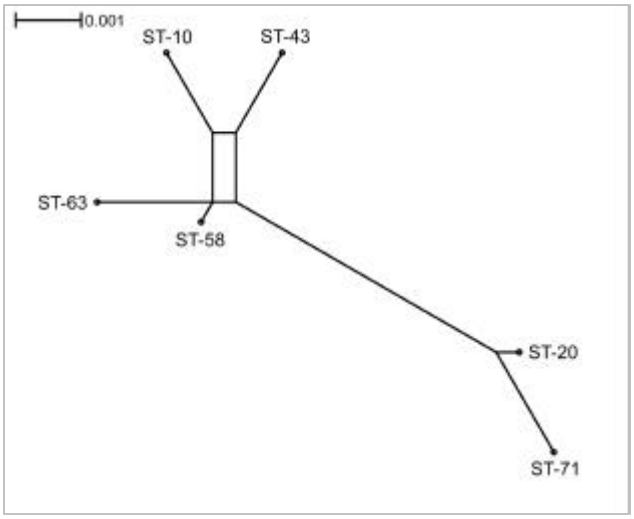


Figure 3.15 SplitsTree graph for the Thuringiensis lineage of clade 2 (bar = number of nucleotide changes per site)

Number of nucleotide changes in the Tolworthi lineage							
ST change	***glpF***	***gmk***	***ilvD***	***pta***	***pur***	***pycA***	***tpi***
ST 22 to ST 4	2				1		2
ST 22 to ST 14	1		1	1	1	5	1
ST 22 to ST 17			16			1	4
ST 22 to ST 19			1		1		1
ST 22 to ST 24			1	2	1		1
ST 22 to ST 29	3			3	1	4	2
ST 22 to ST 34	2		15	2	6	4	2
ST 22 to ST 46	2		15	3	1	4	2
ST 22 to ST 48	2		14			2	1
ST 22 to ST 50	2	1	1	1	2		1
ST 22 to ST 52	5	2	15	1		4	1
ST 22 to ST 53	2	1	16	3	2	4	2
ST 22 to ST 64	2		14	2	7	4	3
ST 22 to ST 73	2		1	2			2
ST 22 to ST 86	2		15	2	1	4	2

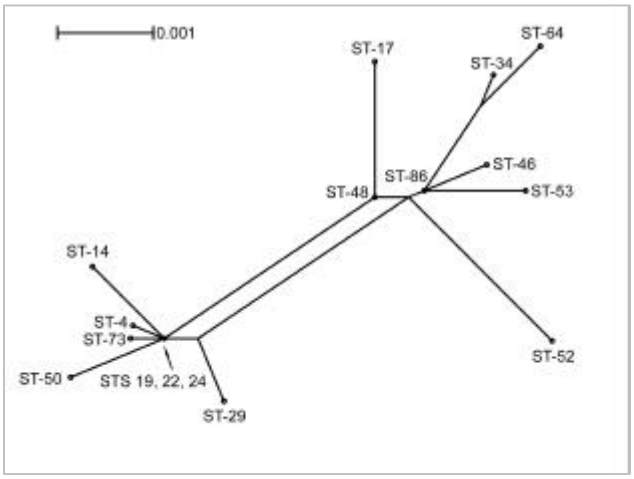


Figure 3.16 SplitsTree graph for the Tolworthi lineage of clade 2 (bar = number of nucleotide changes per site)

Number of nucleotide changes in the Kurstaki lineage							
ST change	***glpF***	***gmk***	***ilvD***	***pta***	***pur***	***pycA***	***tpi***
ST 8 to ST 13	2		4	2			
ST 8 to ST 15	1						1
ST 8 to ST 25	1						
ST 8 to ST 33	2	1	1	3	5	2	
ST 8 to ST 39	4	1	1	2	7	5	
ST 8 to ST 40	5	1	2	2	7	5	
ST 8 to ST 44	4	1	2	2		2	
ST 8 to ST 51	5	1	2	3	3	4	
ST 8 to ST 54	3	1	2	3	8	5	
ST 8 to ST 59	2	1	8	3	5	5	
ST 8 to ST 66	1	1	2	1	1	2	
ST 8 to ST 68	2	1	7	3	5	3	1
ST 8 to ST 84	4	1	4	3	7	3	
ST 8 to ST 85	3	1	2	2		2	

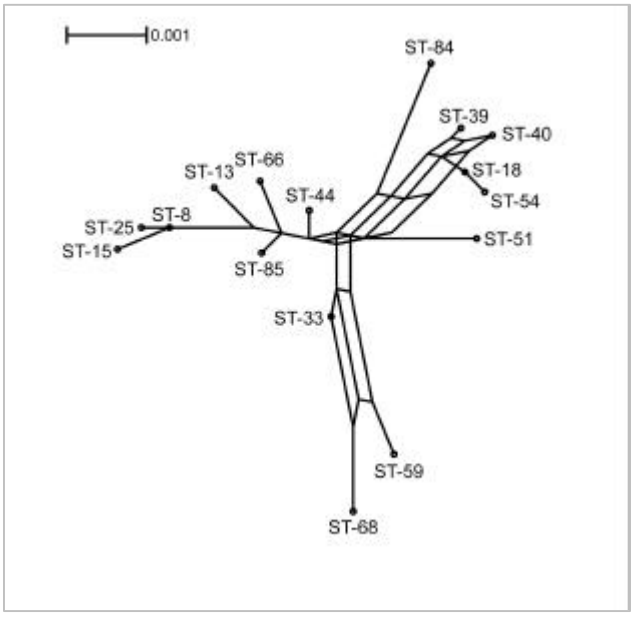


Figure 3.17 SplitsTree graph for the Kurstaki lineage of clade 2 (bar = number of nucleotide changes per site)

Number of nucleotide changes in the 'Others' of Clade 3							
ST change	***glpF***	***gmk***	***ilvD***	***pta***	***pur***	***pycA***	***tpi***
ST 21 to ST 35	3	30			6	2	
ST 21 to ST 36		31	27	5	9	3	2
ST 21 to ST 37		30	34	5	9	3	2
ST 21 to ST 41	3	30	1	6	7	2	
ST 21 to ST 42	5	31	35	4	13	18	2
ST 21 to ST 88		30	1		3		

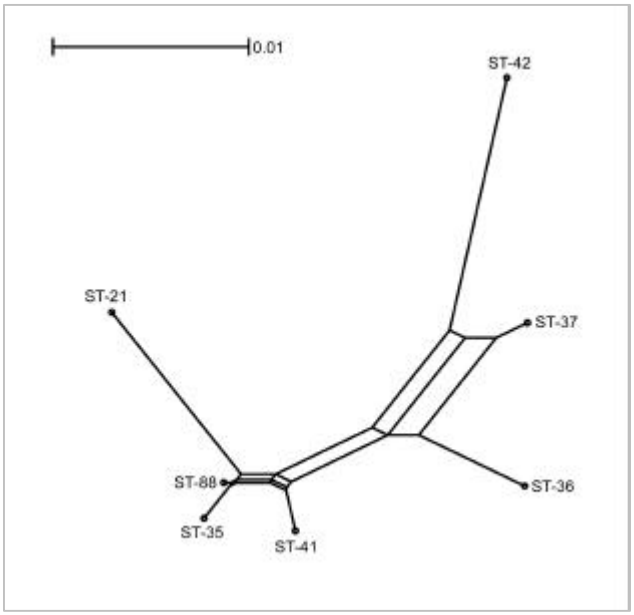


Figure 3.18 SplitsTree graph for the Others - clade 3
(bar = number of nucleotide changes per site)

3.1.6 Further analysis of the seven alleles individually

The phylogenetic analyses presented earlier were carried out on concatenated sequence data. In order to gain some insight into lateral gene transfer among the STs, NJ trees were prepared for each gene (included in appendix E). These trees are quite complex and to simplify the analysis, 21 representative STs were selected from the rooted phylogenetic tree shown in Figure 3.9, and the positions of these 21 STs were identified on each of the seven gene trees.

A reduced NJ tree was produced (Figure 3.19) which showed that the structure of the NJ tree derived from 21 STs was very similar to the earlier NJ tree constructed using all 88 STs (Figure 3.9), in that the three clades and nine lineages were retained, and their relative positions were the same.

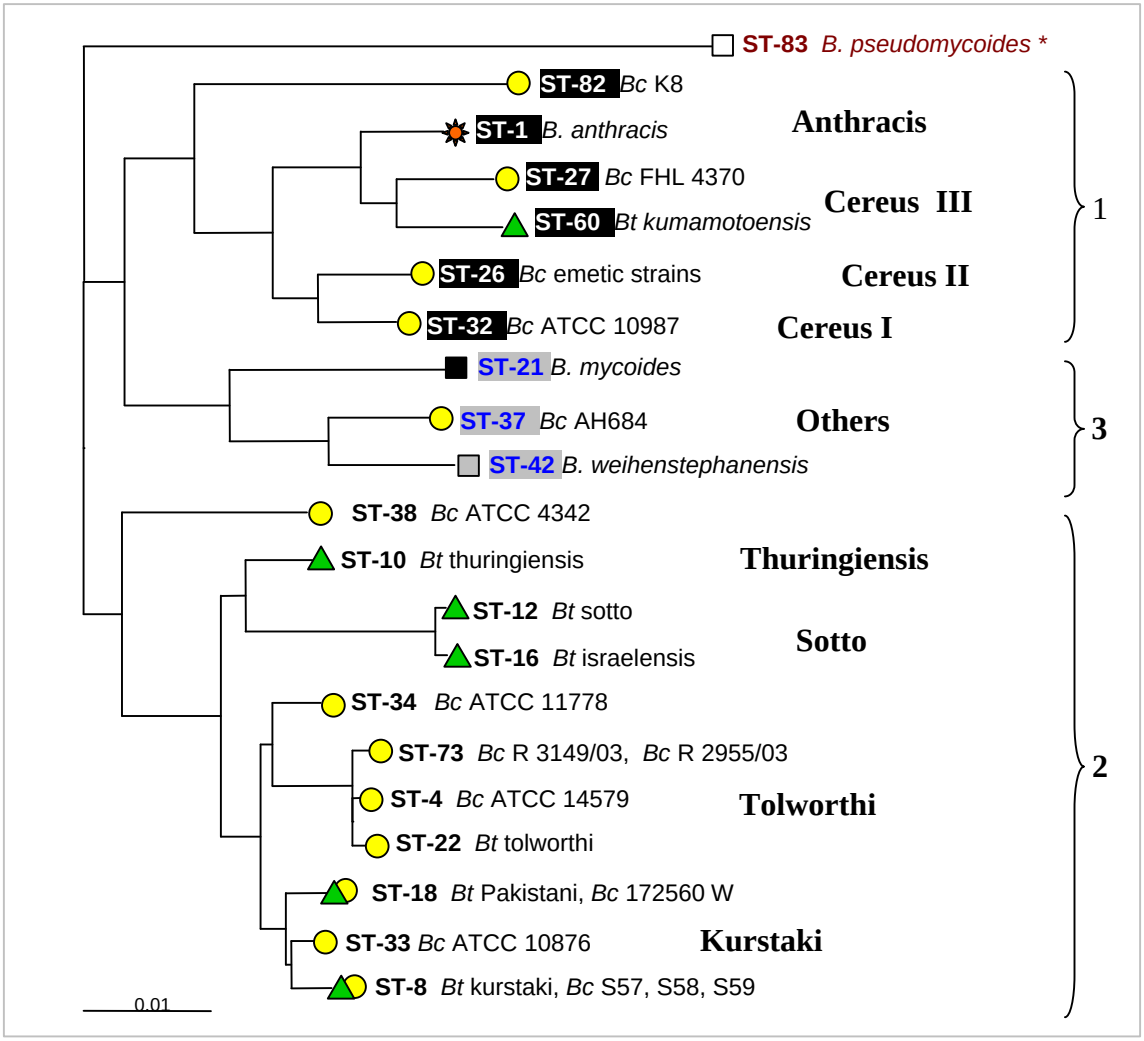


Figure 3.19 Neighbour Joining tree of 21 representative STs, constructed using the concatenated sequence data of 2838 bases and rooted with ST-83 *(B. pseudomycoides).* Key to species symbols as presented in Figures 3.8 and 3.9.

3.1.6.i Neighbour Joining (NJ) Trees of seven alleles with a reduced dataset.

Individual NJ trees were drawn for each of the seven alleles (Figures 3.20 to 3.26). To clarify any structural differences between the trees, annotation has not been applied to the 21 STs. To annotate each gene tree with all the STs was too congested, making the figures difficult to read, because some alleles were present in many STs, e.g. allele *glp-15* is found in 14 STs and allele *gmk-8* is common to 17 STs.

glpF (Figure 3.20) The lineages apparent in the concatenated tree were not clearly defined in the *glpF* tree. The Anthracis lineage was close to the 'outliers', and Cereus I, Kurstaki and ST-38 (ATCC 4342) were not distinguished, suggesting sharing of this allele between these STs. However, the 'others' lineage was retained.

gmk (Figure 3.21) The structure of the *gmk* tree was in agreement with the tree for the concatenated data, apart from ST-21 from the 'others' moving to clade 2.

ilvD (Figure 3.22) The structure of the *ilvD* tree was different from the concatenated tree. ST-8, *Bt* kurstaki was close to *B. pseudomycoides* (the outlier). ST-21, *B. mycoides* appeared in the same cluster as Anthracis / Cereus revealing sharing of this allele by these STs.

pta (Figure 3.23) The structure of the *pta* tree agreed with the concatenated tree showing that it appears to be evolving in line with the organisms and is among the least perturbed by recombination.

pur (Figure 3.24) The *pur* tree was broadly similar to the concatenated tree, with the exception that the Anthracis lineage appeared close to ST-38, *B. cereus* ATCC 4342. ST-18 (a *Bc* and a *Bt* strain) was recovered in the same group as clade 1 strains.

pycA (Figure 3.25) The *pycA* tree was similar to the tree for the concatenated data, but as in the pur tree, ST-38 was found in clade 1.

tpi (Figure 3.26) The *tpi* tree was generally similar to the concatenated tree, with the exception that ST-12 *Bt* sotto and ST-16 *Bt* israelensis formed a group with the 'others'.

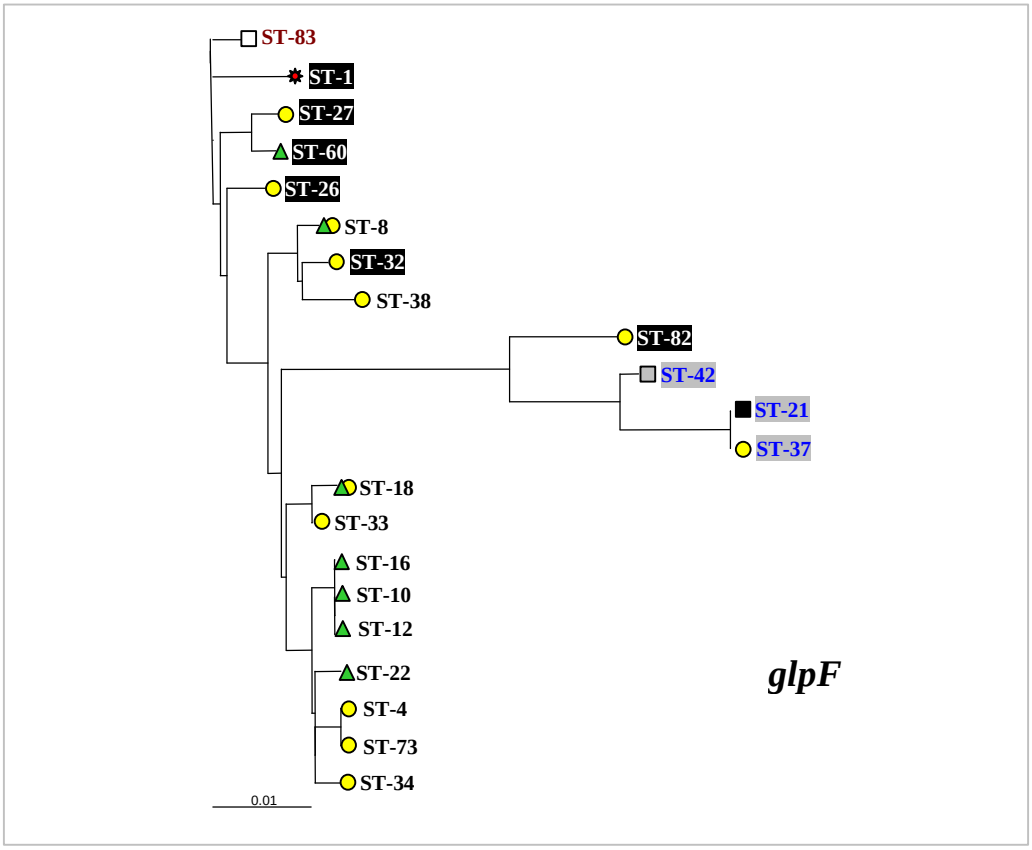


Figure 3.20 Neighbour Joining tree for *glpF* alleles from 21 representative STs, rooted with ST-83, *B. pseudomycoides*.

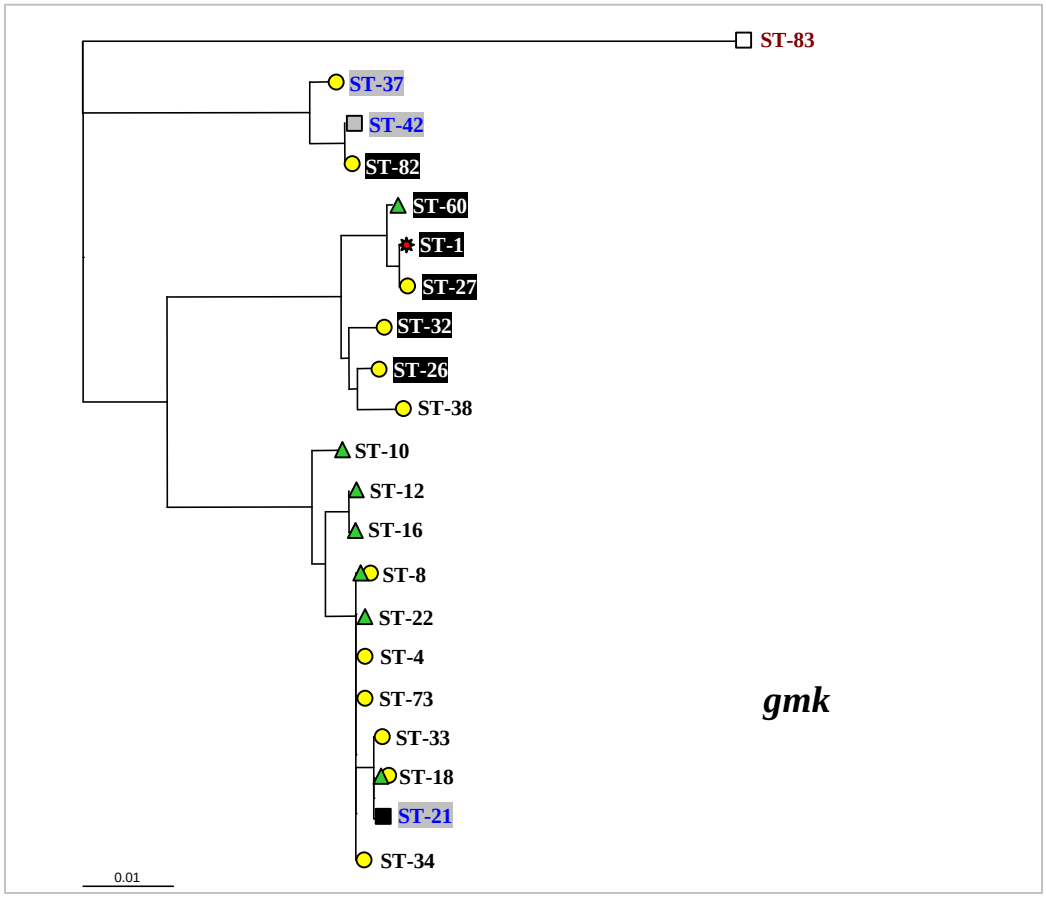


Figure 3.21 Neighbour Joining tree for *gmk* alleles from 21 representative STs, rooted with ST-83, *B. pseudomycoides*

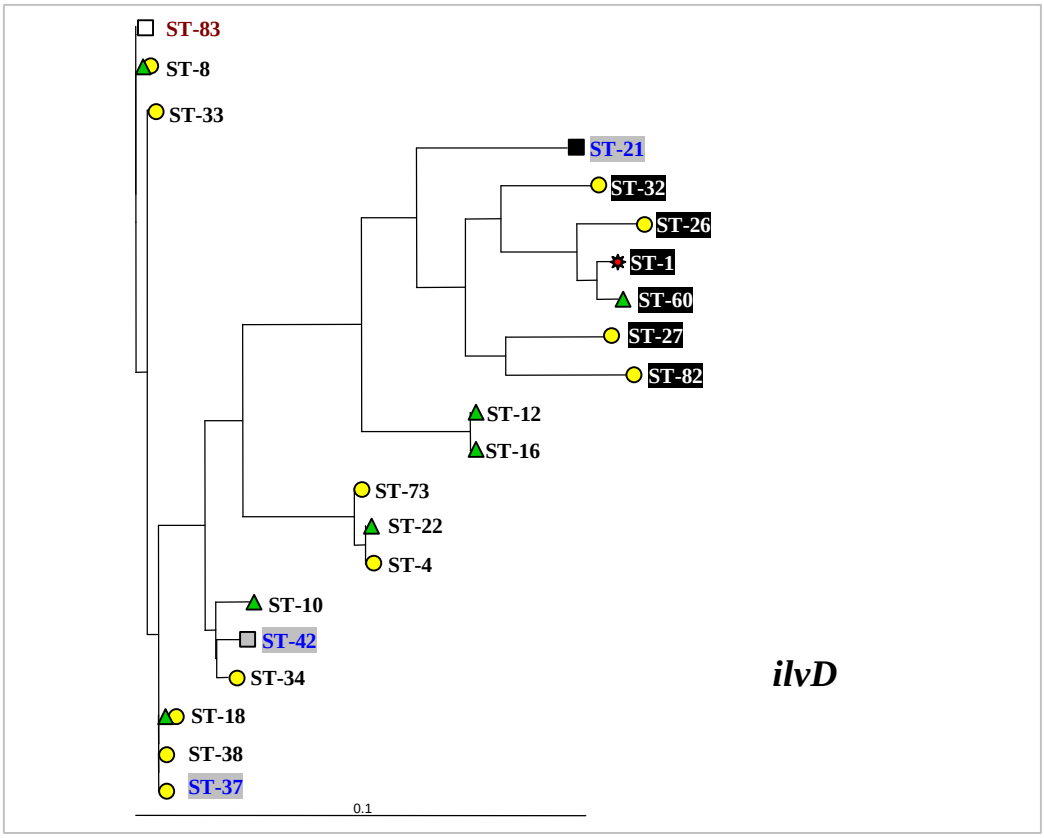


Figure 3.22 Neighbour Joining tree for *ilvD* alleles from 21 representative STs, rooted with ST-83, *B. pseudomycoides*

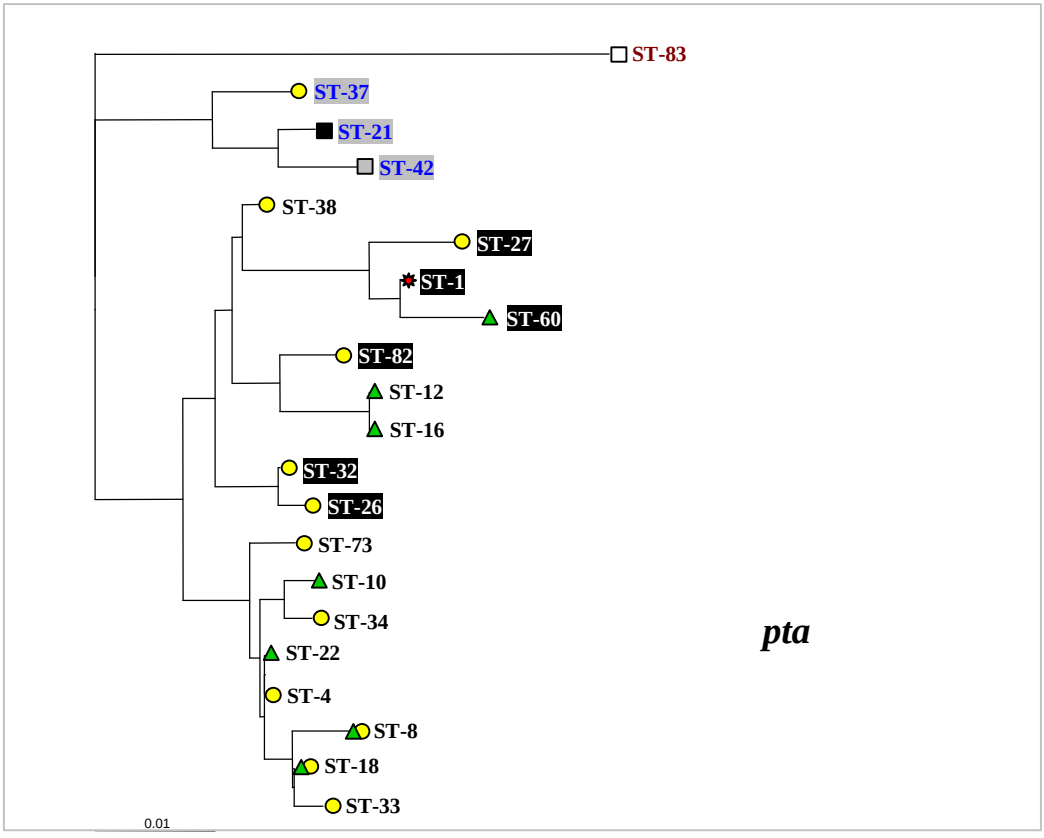


Figure 3.23 Neighbour Joining tree for *pta* alleles from 21 representative STs, rooted with ST-83, *B. pseudomycoides*

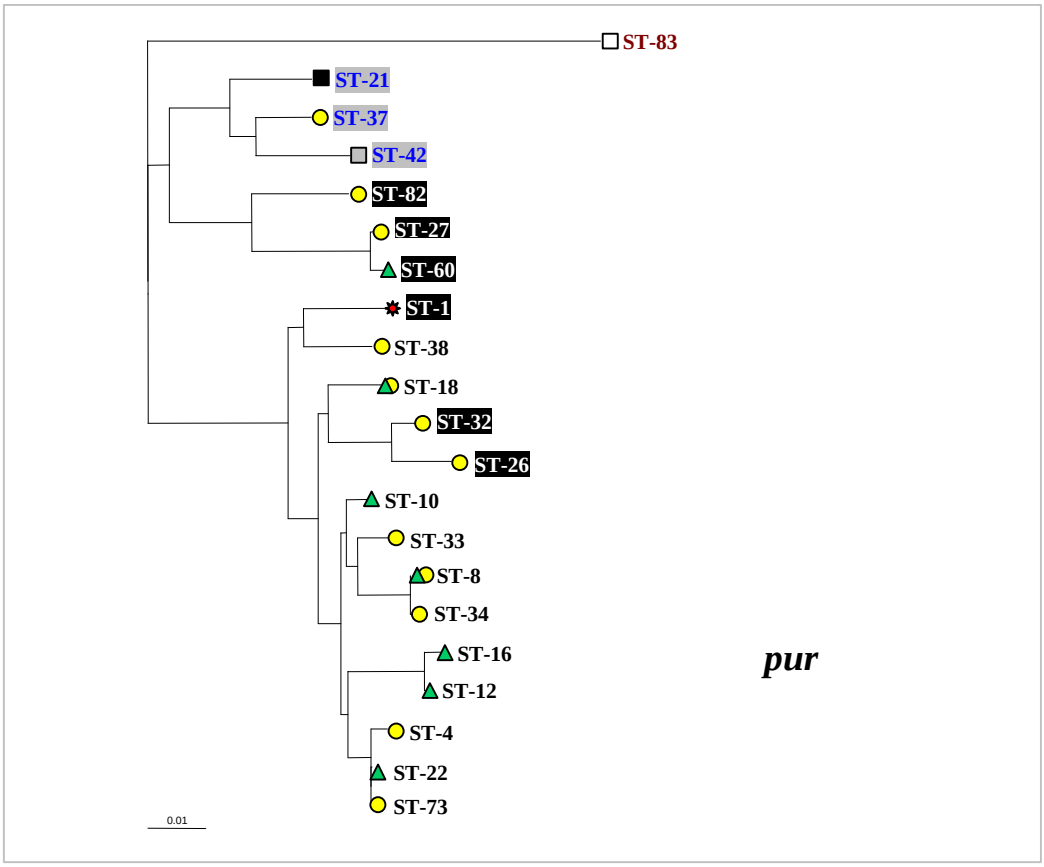


Figure 3.24 Neighbour Joining tree for *pur* alleles from 21 representative STs, rooted with ST-83, *B. pseudomycoides*

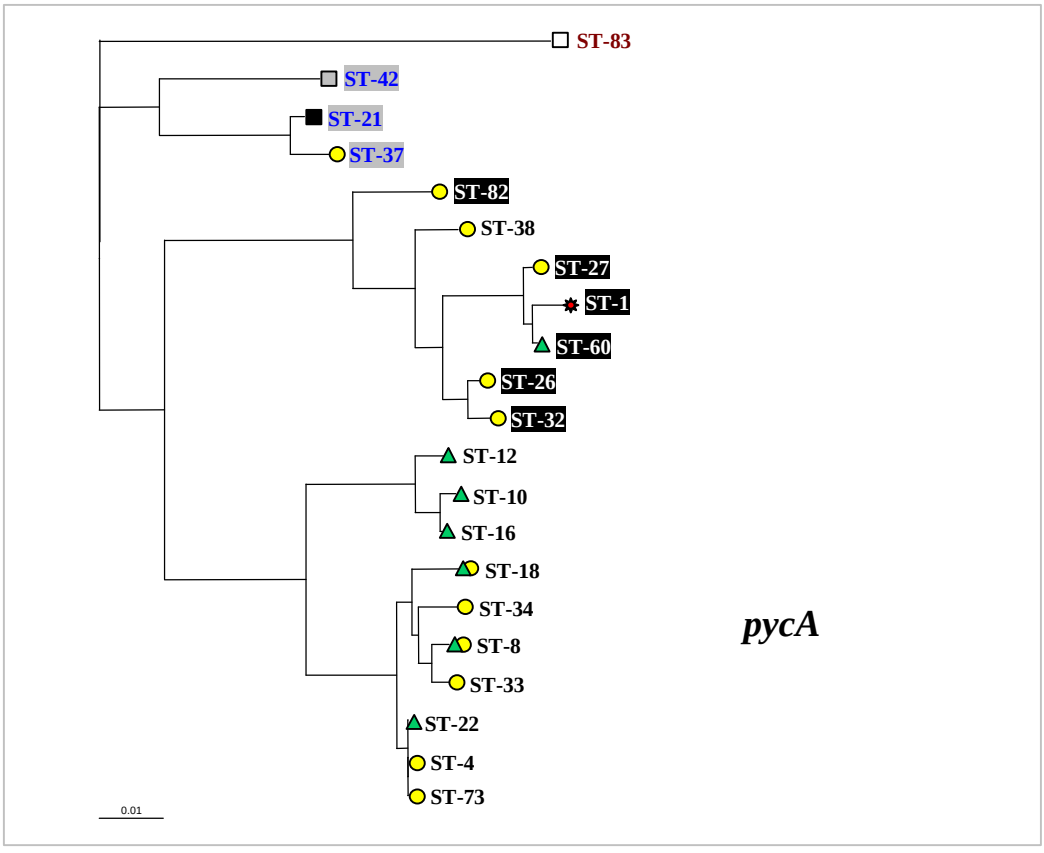


Figure 3.25 Neighbour Joining tree for *pycA* alleles from 21 representative STs, rooted with ST-83, *B. pseudomycoides*

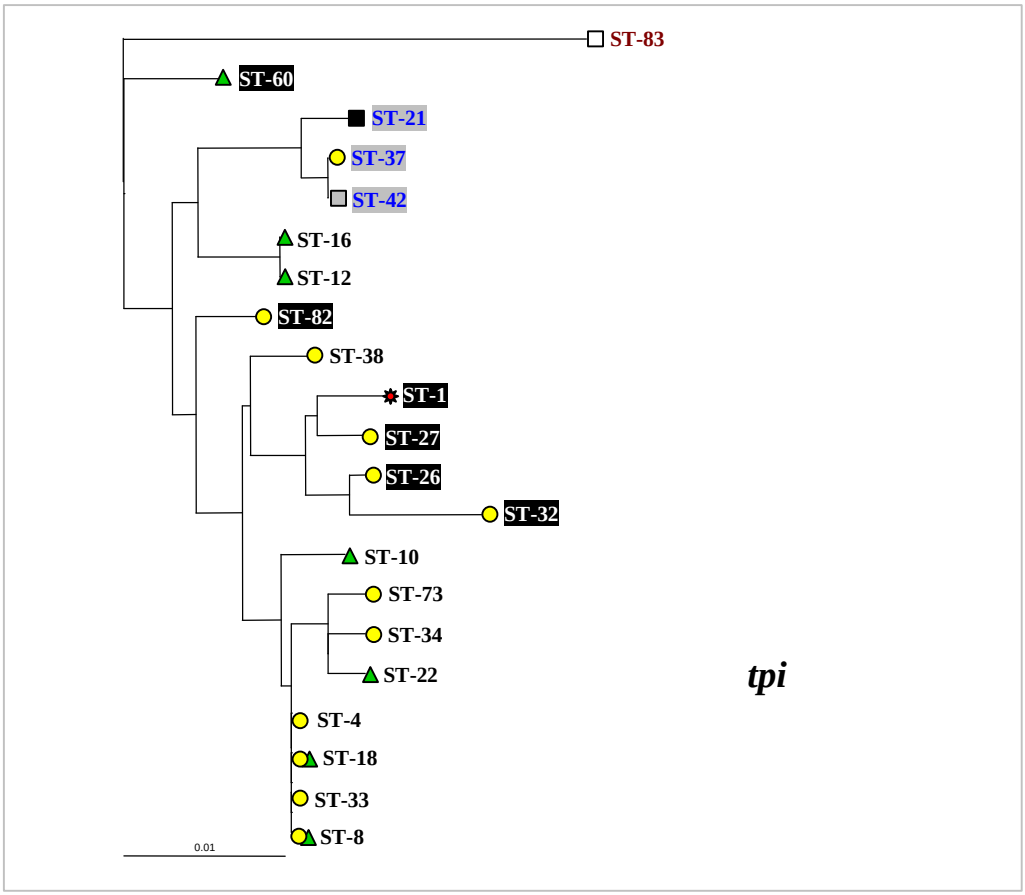


Figure 3.26 Neighbour Joining tree for *tpi* alleles from 21 representative STs, rooted with ST-83, *B. pseudomycoides*

In summary, the *glpF* and *ilvD* trees were found to be quite different from the concatenated tree indicating that these genes do not reflect the organism trees and are the most influenced by lateral gene transfer. The other five gene trees were reasonably similar to the concatenated tree.

3.1.6.ii SplitsTrees of seven alleles with the reduced ST set

SplitsTree graphs were produced for the 21 representative STs, as defined in Figure 3.19, for the concatenated sequences and for each allele, as shown in Figures 3.27 to 3.34.

SplitsTrees for 21 representative STs

The SplitsTree graph of the 21 STs (Figure 3.27) confirmed that the data split naturally into three distinct clades and that ST-83 (*B. pseudomycoides*) is a remote outlier. The SplitsTree graph for individual loci did not always show this division: -

The ***glpF*** SplitsTree graph (Figure 3.28) showed that the lineage called 'others' comprising ST-21, ST-37 and ST-42 was distant from the other STs. Interestingly, for this locus, ST-83 (the outlier) was very close to the Cereus III and Anthracis lineages (see also NJ tree, Fig 3.20) suggesting a recent common origin to this gene. For this gene, ST-82 was recovered in association with the 'others' group.

In the ***gmk*** SplitsTree (Figure 3.29) ST-83 was far removed from the other STs, in accordance with the concatenated sequence analysis. ST-21 (*B. mycoides*) was not recovered with the 'others' as expected, but was close to clade 2 STs, from the Kurstaki and Sotto lineages. As found in *glpF*, ST-82 appears in the 'others' group.

The ***ilvD*** SplitsTree graph (Figure 3.30) failed to resolve the clades as compact groups. Clade 1 STs were clustered together at the left-hand side of the graph, and clade 2 to the right. The 'others' did not group together and ST-83 shared the ilv allele with ST-8 (*Bt* kurstaki) suggesting the lateral transfer of this allele.

The ***pta, pur, pycA*** SplitsTree graphs (Figures 3.31, 3.32, 3.33) were largely in agreement with the concatenated graph, in that ST-83 was on its own, the 'others' grouped together, and clades 1 and 2 were on the left and right hand sides respectively.

In the ***tpi*** SplitsTree graph (Figure 3.34) ST-83 was a definite outlier again, but clades 1, 2 and 3 were close together.

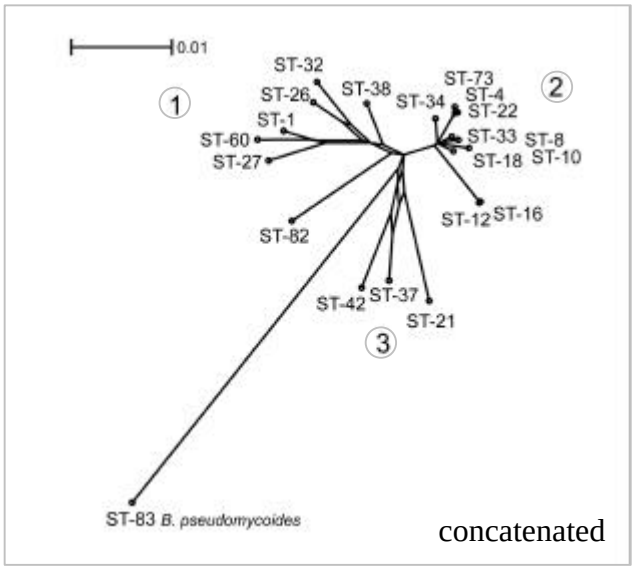


Figure 3.27 SplitsTree graph of 21 representative STs. Concatenated sequences, showing clades 1, 2 and 3.

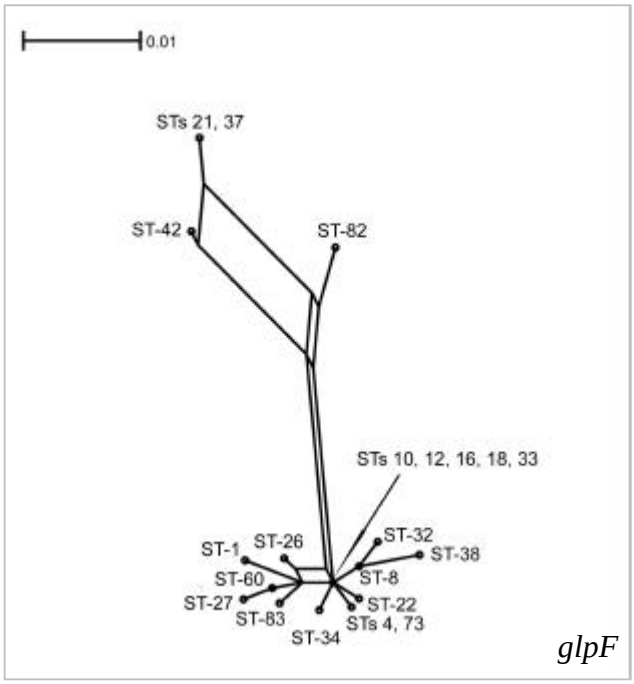


Figure 3.28 SplitsTree graph for *glpF* allele. (21 representative STs).

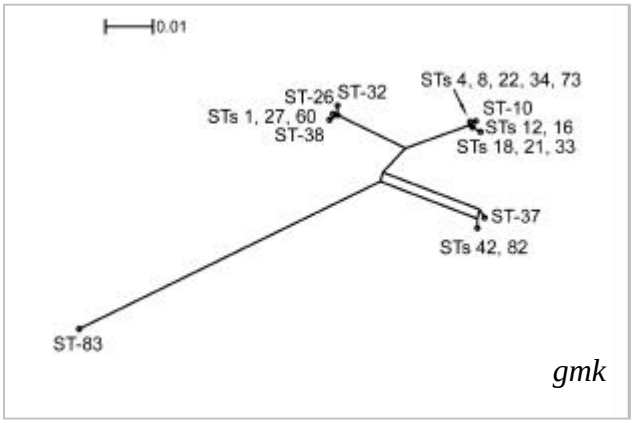
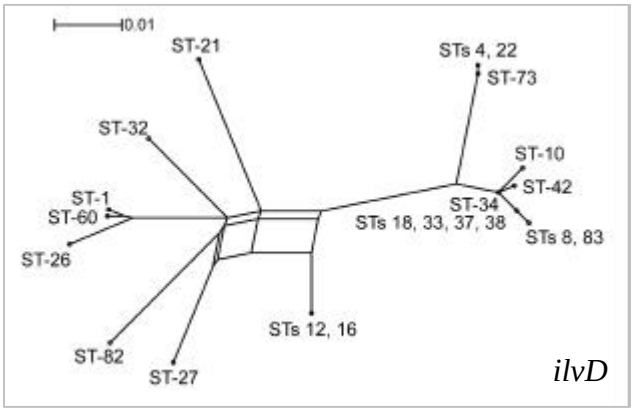


Figure 3.29 SplitsTree graph for *gmk* allele. (21 representative STs).



'igure 3.30 SplitsTree graph for *ilvD* allele. (21 representative STs).

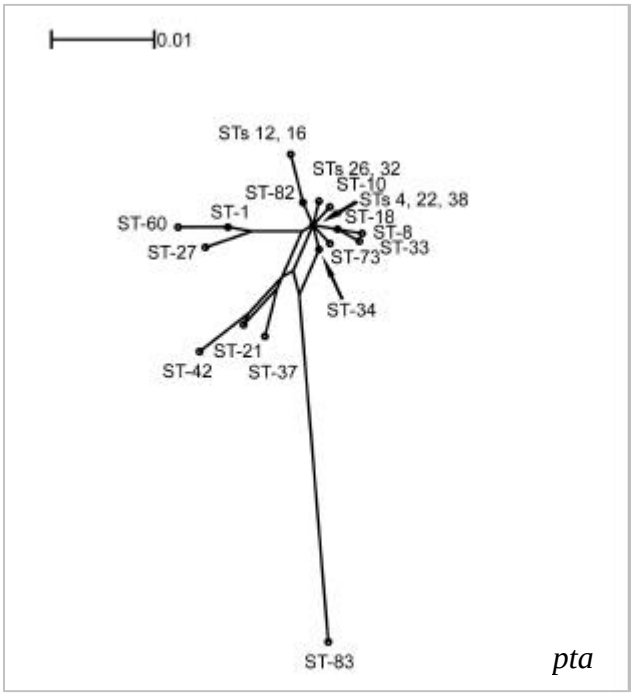


Figure 3.31 SplitsTree graph for *pta* allele. (21 representative STs).

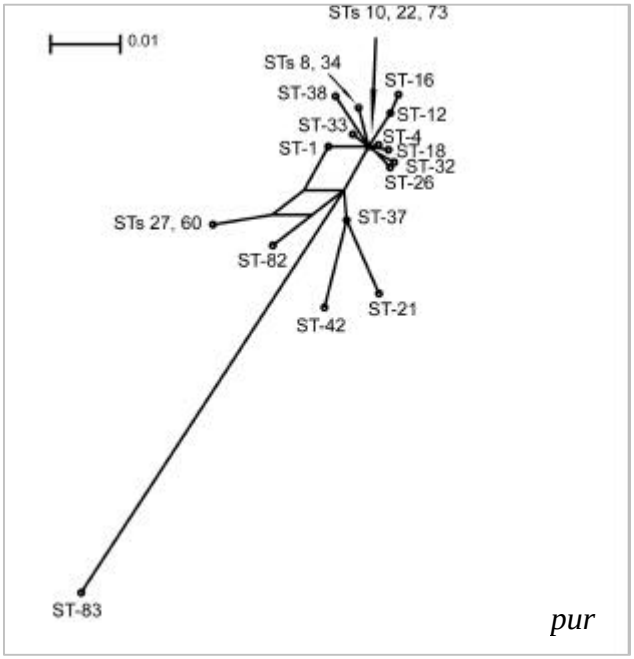


Figure 3.32 SplitsTree graph for *pur* allele. (21 representative STs).

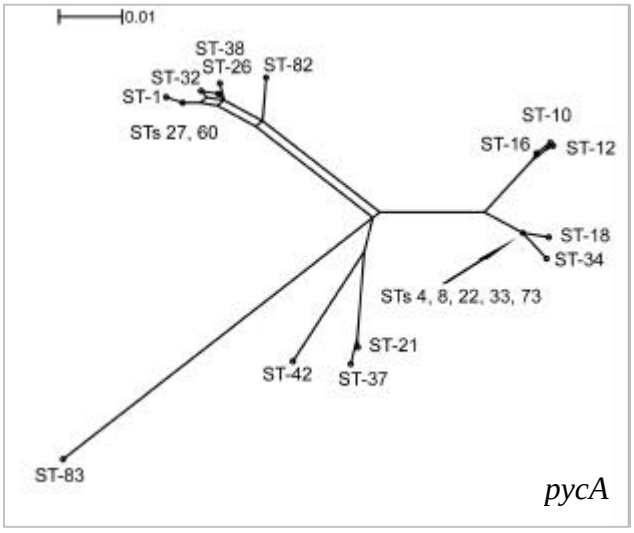


Figure 3.33 SplitsTree graph for *pycA* allele. (21 representative STs).

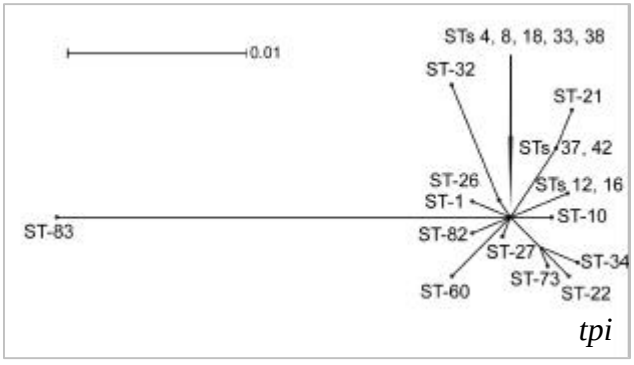


Figure 3.34 SplitsTree graph for *tpi* allele. (21 representative STs).

3.1.7 Further tests

3.1.7.i Maximum likelihood randomization test

A test to compare the structures of the concatenated tree and the seven individual gene trees was carried out on a dataset with 105 strains (59 STs). This work was carried out by E. Holmes at Oxford University, and is discussed in the paper submitted with this thesis ('Population Structure and Evolution of the *Bacillus cereus* group', Priest *et al* (2004)). E. Holmes found that 'while the topologies of the eight trees had different likelihoods, indicating that the signals present in these data were not completely congruent and therefore that the population was not entirely clonal, they were far more similar than would be expected by chance alone. As such, there is a clear signal of phylogenetic history present in these MLST data so that they can be used to reconstruct an evolutionary history of the *B. cereus* group'.

The dataset used in the paper referred to above was smaller than the dataset examined in this thesis, 105 strains (59 STs) were used for the paper and this was extended to 146 strains (88 STs) for this thesis. The graphs presented in the paper support the contention that the trees from each gene and the concatenated tree are more similar to each other than they are to trees generated randomly.

3.1.7.ii Index of association

Analysis of the Index of Association (I_A) was carried out using the START program (Jolley, 2000). I_A measures the amount of recombination among a set of DNA sequences. For a panmictic population, at linkage equilibrium, the value of I_A will be zero. Clonal populations will differ significantly from zero.

$$\mathbf{I_A = VO/VE - 1}$$

VO is the observed variance. VE is the expected variance if the alleles present at different loci in an individual are independent (no linkage disequilibrium). I_A was calculated for all 146 strains, and for the 88 STs.

146 strains	**88 STs**
VO = 1.808, VE = 0.471	VO = 0.805, VE = 0.294
I_A =VO/VE-1 = **2.841**	I_A =VO/VE-1 = **1.737**

The difference in the value of I_A calculated using all strains and calculated using STs only may be influenced by the method of sampling, but indicates that the population structure may be weakly clonal. It has been suggested that such a difference in I_A values is characteristic of an epidemic type of population (Holmes *et al*,1999).

3.1.8 Final Lineage Assignment

From the analysis of the MLST data presented in sections 3.1.2 to 3.1.7 it is apparent that the phylogenetic structure of the *Bacillus cereus* group, as determined from the 146 strains studied here, does not separate into species, as currently defined, but into three clades, which can be sub-divide into nine lineages.

The details of the 146 strains studied, their STs and final allocation to clades and lineages are shown in Table 3.14.

Table 3.14 Strains studied and their allocation to lineages:-

Sequence type	Strain	Country of Origin	Year isolated	Original designation[a]	Ref. /Source[b]
Clade 1 Anthracis					
ST-1	Ames	USA	1982	*B. anthracis*	TDR
ST-1	Ames[C]	USA		*B. anthracis*	TDR
ST-1	K0610/A0034	China		*B. anthracis*	PK
ST-1	K4834/A0039	Australia	1994	*B. anthracis*	PK
ST-1	K1340/A0062	Poland	1962	*B. anthracis*	PK
ST-1	K1694/A0462	USA	1932	*B. anthracis*	PK
ST-1	K5135/A0463	Pakistan	1978	*B. anthracis*	PK
ST-1	K4596/A0488	UK	1997	*B. anthracis*	PK
ST-2	K3700/A0267	USA	1937	*B. anthracis*	PK
ST-3	K2478/A0102	Mozambique	1944	*B. anthracis*	PK
ST-3	K2762/A0465	France	1997	*B. anthracis*	PK
Clade 1 Cereus I					
ST-5	m1545	Brazil	1987	*B. cereus*	MADM
ST-6	m1564	Brazil	1987	*B. cereus*	MADM
ST-7	M21	Finland	1998	*B. cereus*	MADM
ST-32	ATCC 10987	Canada	1930	*B. cereus*	ATCC
Clade 1 Cereus II (emetic)					
ST-26	F4810/72	USA	1972	*B. cereus*	PCBT
ST-26	F4746/77		1977	*B. cereus*	FGP
ST-26	S710	UK	1979	*B. cereus*	FGP
ST-26	F3080B/87	UK	1987	*B. cereus*	TSP

Sequence type	Strain	Country of Origin	Year isolated	Original designation[a]	Ref. /Source[b]
ST-26	F3942/87	UK	1987	*B. cereus*	TSP
ST-31	S366	North Sea	1988	*B. cereus*	FGP
ST-45	m1293	Brazil	1987	*B. cereus*	MADM
ST-47	m1576	Brazil	1987	*B. cereus*	MADM
ST-81	S78		1988	*B. cereus*	FGP
Clade1 Cereus III					
ST-11	2002734341	USA	2003	*B. cereus*	CDC
ST-27	F4370/75	UK	1975	*B. cereus*	FGP
ST-57	T10024	Pakistan	1975	*Bt* darmstadiensis	IP
ST-60	T18004	Iraq	1984	*Bt* kumamotoensis	IP
ST-62	2002734374	USA	2003	*B. cereus*	CDC
ST-75	R 3039/03	UK	2003	*B. cereus*	MB
ST-76	191560 K	UK	2003	*B. cereus*	MB
ST-77	168287 M	UK	2003	*B. cereus*	MB
Clade 2 Kurstaki					
ST-8	S57	USA	1975	*B. cereus*	BA
ST-8	S58	USA	1975	*B. cereus*	BA
ST-8	S59	USA	1975	*B. cereus*	BA
ST-8	T03a001	France	1961	*Bt* kurstaki	IP
ST-8	T03a075	Iraq	1976	*Bt* kurstaki	IP
ST-8	T03a172	Pakistan	1982	*Bt* kurstaki	IP
ST-8	T03a287	Kenya	1988	*Bt* kurstaki	IP
ST-8	T03a361	Australia	1990	*Bt* kurstaki	IP

Sequence type	Strain	Country of Origin	Year isolated	Original designation[a]	Ref. /Source[b]
ST-13	T04b001	Kenya	1962	*Bt* kenyae	IP
ST-13	T04b012	Bulgaria	1974	*Bt* kenyae	IP
ST-13	T04b054	Iraq	1986	*Bt* kenyae	IP
ST-13	T04b060	Iraq	1987	*Bt* kenyae	IP
ST-13	T04b073	Chile	1993	*Bt* kenyae	IP
ST-15	T07033	Japan	1975	*Bt* aizawai	IP
ST-15	T07058	France	1983	*Bt* aizawa*i*	IP
ST-15	T07180	Spain	1992	*Bt* aizawai	IP
ST-18	T13028	Chile	1993	*Bt* pakistani	IP
ST-18	172560 W	UK	2003	*B. cereus*	MB
ST-25	T05005	USA	1964	*Bt* pakistani	IP
ST-25	T05033	USA	1975	*Bt* pakistani	IP
ST-25	T05144	France	1985	*Bt* pakistani	IP
ST-33	ATCC 10876			*B. cereus*	ATCC
ST-39	SPS 2		1999	*B. cereus*	TSP
ST-40	TSP 11		1999	*B. cereus*	TSP
ST-44	m1292	Brazil	1987	*B. cereus*	MADM
ST-51	T05a015	USA	1977	*Bt* canadensis	IP
ST-54	T07196	Brazil	1993	*Bt* aizawai	IP
ST-59	T18001	Japan	1980	*Bt* kumamotoensis	IP
ST-59	T18002	USA	1980	*Bt* kumamotoensis	IP
ST-66	m1552	Brazil	1987	*B. cereus*	MADM
ST-68	T09016	Mexico	1988	*Bt* tolworthi	IP
ST-68	T09043	Thailand	1994	*Bt* tolworthi	IP

Sequence type	Strain	Country of Origin	Year isolated	Original designation[a]	Ref. /Source[b]
ST-84	2002734432	USA	2003	*B. cereus*	CDC
ST-85	2002734433	USA	2004	*B. cereus*	CDC
Clade 2 Sotto					
ST-12	T04002	Canada	1965	*Bt* sotto	IP
ST-12	T04016	Pakistan	1980	*Bt* sotto	IP
ST-12	T04024	Pakistan	1981	*Bt* sotto	IP
ST-12	T015001	USA	1983	*Bt* dakota	IP
ST-16	CCCT 2259	Brazil	1993	*Bt* israelensis	DAK
ST-16	T08025	France	1988	*Bt* morrisoni	IP
ST-16	T05a030	Iraq	1987	*Bt* canadensis	IP
ST-23	T08001	USA	1963	*Bt* morrisoni	IP
ST-23	T08009	USA	1979	*Bt* morrisoni	IP
ST-23	T08012	Pakistan	1980	*Bt* morrisoni	IP
ST-23	T08023	Brazil	1987	*Bt* morrisoni	IP
ST-23	T08031	Brazil	1991	*Bt* morrisoni	IP
ST-23	4Btt	USA		*Bt* tenebrionis	AY
ST-49	T04236	Indonsesia	1991	*Bt* sotto	IP
ST-55	T10016	USA	1982	*Bt* darmstadiensis	IP
ST-56	T10001	Germany	1967	*Bt* darmstadiensis	IP
ST-56	T10003	Germany	1967	*Bt* darmstadiensis	IP
ST-56	T10017	USA	1982	*Bt* darmstadiensis	IP
ST-56	T10018	Japan	1982	*Bt* darmstadiensis	IP
ST-61	T05125	S. Korea	1993	*Bt* galleriae	IP
ST-65	T14531	Vietnam	1998	*Bt* israelensis	IP
ST-67	T08121	Peru	1999	*Bt* morrisoni	IP

Sequence type	Strain	Country of Origin	Year isolated	Original designation[a]	Ref. /Source[b]
ST-69	T08016	Germany	1982	*Bt* morrisoni	IP
ST-70	T08051	Canada	1992	*Bt* morrisoni	IP
Clade 2 Thuringiensis					
ST-10	T01001	Canada	1958	*Bt* thuringiensis	IP
ST-10	T01015	Bulgaria	1962	*Bt* thuringiensis	IP
ST-10	T01022	USA	1964	*Bt* thuringiensis	IP
ST-10	T01058	Switzerland	1964	*Bt* thuringiensis	IP
ST-10	T01326	Chile	1993	*Bt* thuringiensis	IP
ST-20	WSBC 10312	Thailand	1999	*B. cereus*	BMP
ST-43	m1278	Brazil	1987	*B .cereus*	MADM
ST-58	T15006	South Korea	1993	*Bt* dakota	IP
ST-63	T08019	Mexico	1985	*Bt* morrisoni	IP
ST-71	T09020	Italy	1990	*Bt* tolworthi	IP
Clade 2 Tolworthi					
ST-4	ATCC 14579^{T}	USA	1916	*B. cereus*	ATCC
ST-14	T006010	Pakistan	1983	*Bt* entomocidus	IP
ST-14	T006010	Pakistan	1983	*Bt* entomocidus	IP
ST-17	T13001	Pakistan	1976	*Bt* pakistani	IP
ST-17	T13004	Pakistan	1980	*Bt* pakistani	IP
ST-19	WSBC 10249	Denmark	1999	*B. cereus*	BMP
ST-19	m1280	Brazil	1987	*B. cereus*	MADM
ST-22	T09010	USA	1979	*Bt* tolworthi	IP
ST-22	T09011	Iraq	1987	*Bt* tolworthi	IP
ST-22	T09024	Indonesia	1991	*Bt* tolworthi	IP
ST-22	T09034	Brazil	1992	*Bt* tolworthi	IP

Sequence type	Strain	Country of Origin	Year isolated	Original designation[a]	Ref. /Source[b]
ST-22	T09052	Canada	1999	*Bt* tolworthi	IP
ST-24	NCIB 6349		1988	*B. cereus*	FGP
ST-24	Cal3	Finland	1998	*B. cereus*	TSP
ST-24	WSBC 10028	Germany	1999	*B. cereus*	BMP
ST-29	S86		1988	*B. cereus*	FGP
ST-34	ATCC 11778			*B. cereus*	ATCC
ST-46	m1550	Brazil	1987	*B. cereus*	MADM
ST-48	T01246	Iraq	1984	*Bt* thuringiensis	IP
ST-50	T05a001	Canada	1968	*Bt* canadensis	IP
ST-52	T05a019	Pakistan	1980	*Bt* canadensis	IP
ST-53	T07146	Indonesia	1991	*Bt* aizawai	IP
ST-64	T08003	Canada	1964	*Bt* morrisoni	IP
ST-73	R 2955/03	UK	2003	*B. cereus*	MB
ST-73	R 3149/03	UK	2003	*B. cereus*	MB
ST-86	TSP 10		1999	*B. cereus*	TSP
Others					
ST-21	WSBC 10277	Germany	1999	*B. mycoides*	BMP
ST-35	AH621	Norway	1998	*B. cereus*	EH
ST-36	AH647	Norway	1998	*B. cereus*	EH
ST-37	AH684	Norway	1998	*B. cereus*	EH
ST-41	WSBC 10202	Germany	1999	*B. weihenstephanensis*	BMP
ST-42	WSBC 10364	Germany	1999	*B. weihenstephanensis*	BMP
ST-88	STR4	UK	2004	*B. mycoides*	MB

Sequence type	Strain	Country of Origin	Year isolated	Original designation[a]	Ref. /Source[b]
Outliers					
ST-83	WS 3118	USA	1999	*B. pseudomycoides*	BMP
ST-87	WSBC 10360	Germany	1999	*B. mycoides*	BMP
Unassigned					
ST-9	NCTC 6474	UK	1988	*B. cereus*	FGP
ST-28	F4431/3	Indonesia	1973	*B. cereus*	FGP
ST-30	S363	North Sea	1988	*B. cereus*	FGP
ST-38	ATCC 4342	USA	1900	*B. cereus*	ATCC
ST-72	R 3238/03	UK	2003	*B. cereus*	MB
ST-74	R 3098/03	UK	2003	*B. cereus*	MB
ST-78	2002734342	USA	2003	*B. cereus*	CDC
ST-79	S79		1988	*B. fluorescens*	FGP
ST-80	TSP 8		1999	*B. cereus*	TSP
ST-82	K8		1999	*B. cereus*	TSP

[a] *B. thuringiensis* strains are given serovar designations where known.

[b] Strains labeled IP are from the Collection of *Bacillus thuringiensis* and *Bacillus sphaericus*, Institut Pasteur, Paris; BA from Brian Austin, Heriot Watt University, Edinburgh, MADM from Marilena Aquino de Muro, CABI Biosciences, UK and AY from Alan Yousten, UPISU, Blacksburg, VA, USA.

References for sources of strains used in MLST :-

MB	Barker *et al,*	2005
DAK	Kaji *et al,*	1994
PK	Keim *et al,*	2000
TSP	Pirrtijarvi *et al,*	1999
FGP	Priest *et al,*	1988
BMP	Prüss *et al,*	1999
TR	Read *et al,*	2003
PT	Turnbull *et al,*	1979

3.2 Random Amplified Polymorphic DNA (RAPD)

The technique of amplifying random fragments of genomic and plasmid DNA using a single short primer was used by Brousseau *et al*. (1993) to study a collection of *B. thuringiensis* strains. Strains with similar banding patterns must have common primer binding sites and can be considered to be phylogenetically close relatives. RAPD PCR reactions were carried out using one of the primers from Brousseau's work (0955-03), 5'-CCG AGT CCA-3'. (Materials and Methods, section 2.6). In the early part of this project, RAPD analysis was carried out on a selection of *B. cereus* and *B. thuringiensis* strains to determine the value of this approach to assess genomic diversity (Figures 3.35 to 3.40). Later, RAPD analysis of selected serotypes of *B. thuringiensis* strains was carried out to investigate possible clonal groups. The plan was to combine and compare all of the data. Unfortunately, although all the RAPD analysis worked, comparison was not possible. A reference strain, T37001 *Bt* andalousiensis was included in all the RAPD gels, as this strain usually gave well-spaced bands. RAPDs that were run at the same time, using the same PCR machine, electrophoresis conditions and batch of primer produced comparable bands for the reference strain. Unfortunately, it was not possible to repeat the original banding patterns for the reference strain using a new PCR machine. However, RAPD gels run at the same time and with similar reference bands, could be compared. In this way, RAPDs may be a useful technique to find out if any strains are clonal, but care must be taken to include a known reference strain, to examine all the strains of interest at the same time, and if possible, to electrophorese them on the same gel.

3.2.1 RAPD analysis of a selection of strains from the *B. cereus* group.

The gels shown in Figures 3.35 to 3.40 were prepared in the early part of the project, and an initial inspection shows that there were a great variety of RAPD patterns produced. However, there were some strains that showed similar patterns, indicating possible clonality. In Figure 3.36, *B. cereus* strains F3080B/87, F3942/97 and F4810/72 have the same RAPD patterns and all three strains are known to be emetic toxin-producers. Two *B. mycoides* strains (WSBC 10277 and WSBC 10360) in Figure 3.36 have different RAPD patterns. Additionally, in Figure 3.38 two strains of *Bt* kurstaki (T03a001 and T03a075) have the same banding patterns as two *B. cereus* strains (S57 and S59).

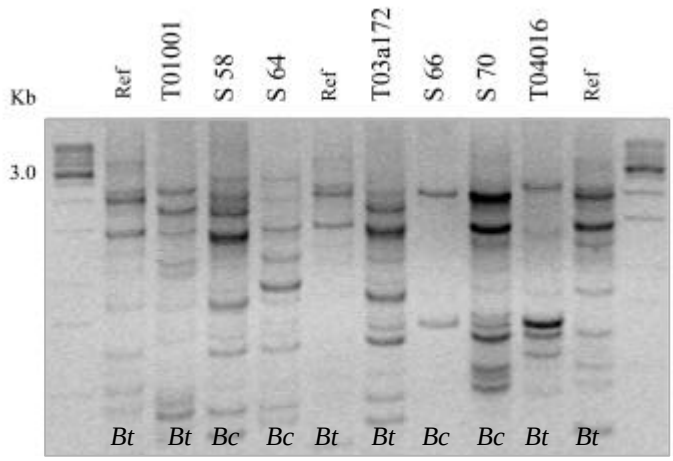


Figure 3.35 RAPD patterns for *Bacillus* strains

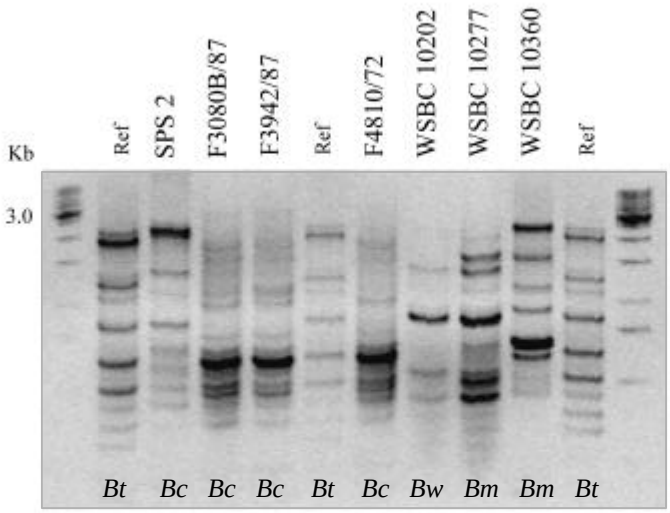


Figure 3.36 RAPD patterns for *Bacillus* strains

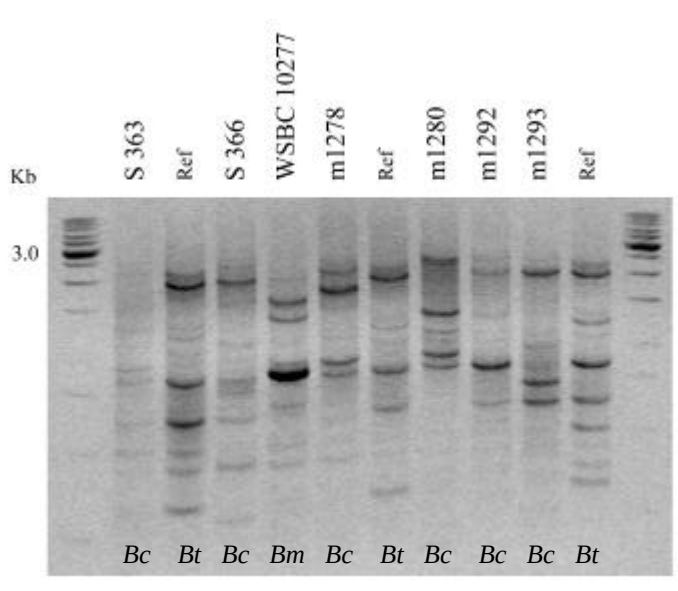


Figure 3.37 RAPD patterns for *Bacillus* strains

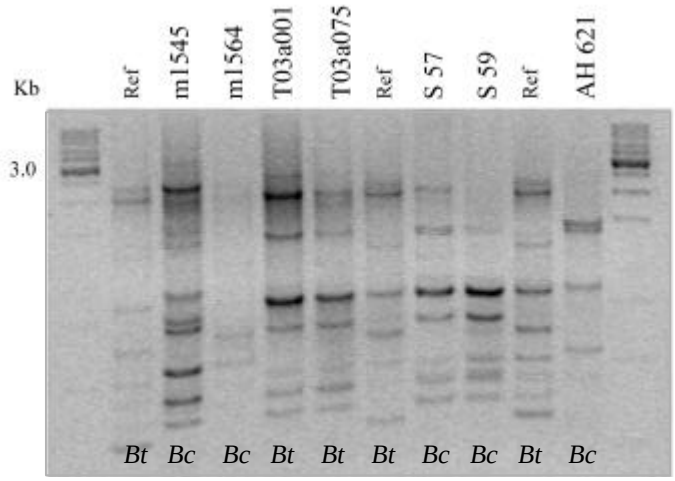


Figure 3.38 RAPD patterns for *Bacillus* strains

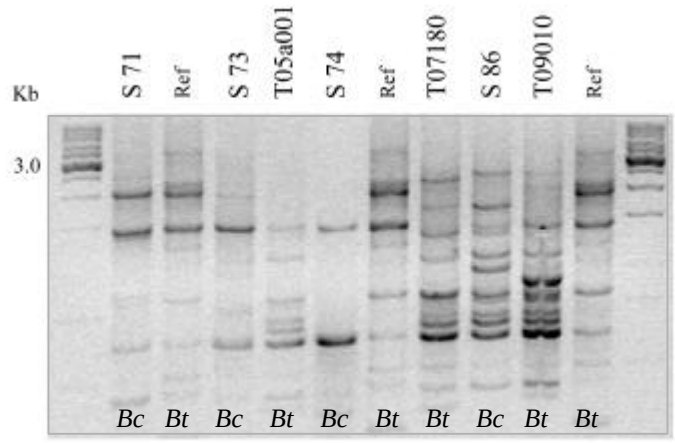


Figure 3.39 RAPD patterns for *Bacillus* strains

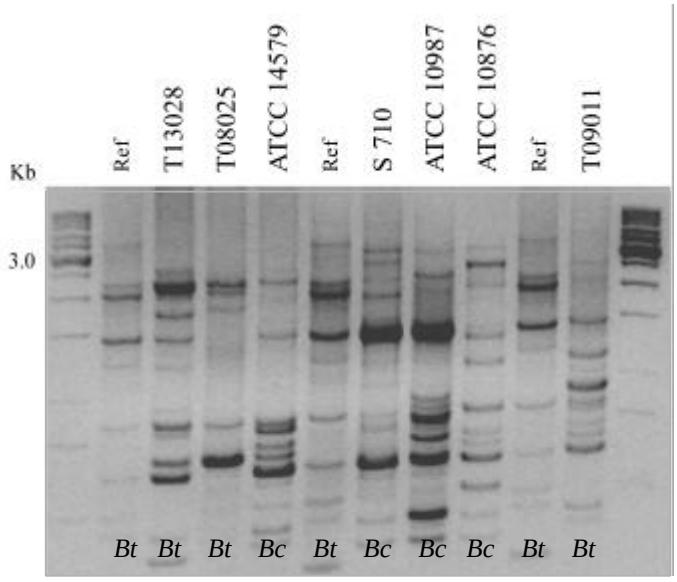


Figure 3.40 RAPD patterns for *Bacillus* strains

RAPD analysis of *B. thuringiensis* strains

MLST analysis had indicated that there may be some clonal groups in Clade 2. The gels shown in Figures 3.41 to 3.47 were electrophoresed in the later part of the project to investigate clonality in the *B. thuringiensis* serotypes of Clade 2. About one hundred strains of *B. thuringiensis* from seven serovars were kindly donated from the collection of the Institut Pasteur. The strains were selected to be as representative as possible for the following serovars: - *Bt* canadensis, *Bt* darmstadiensis, *Bt* israelensis, *Bt* morrisoni , *Bt* sotto, *Bt* thuringiensis and *Bt* tolworthi.

The RAPD gels were electrophoresed under as similar conditions as possible, and all the strains for a particular serotype were examined at the same time. The gels were scrutinised for each serotype and band patterns were assigned numbers for each strain according to similarity and position. For example, in Figure 3.41, *Bt* thuringiensis, strain T01001 was allocated the RAPD number 1. Strain T01246 was completely different from the other strains and was called type 2. Where the RAPD patterns between two strains were very similar, or apparent differences appeared to be due to variation in the intensity of the bands, strains were labelled with an ‘a’ or ‘b’ after the type number.

In general, most serovars showed some degree of clonality, and Table 3.15 gives the rank, from most clonal, with only one deviant strain, to least clonal.

Table 3.15 Clonality of *Bt* serovars, by RAPD analysis

serotype	Most prevalent type	Number of occurrencies	Total number of strains	% clonal
Bt thuringiensis	1_thu	19	20	**95**
Bt israelensis	1_isr	6	7	**86**
Bt morrisoni	1_mor	12	20	**60**
Bt tolworthi	1_tol	10	17	**58**
Bt canadensis	1_can	8	18	**39**
Bt darmstadiensis	1_dar	6	18	**33**
Bt sotto	1_sot, 2_sot	6	18	**33**

Some strains produced identical RAPD patterns to strains from different serovars, for example :-

In Figure 3.42, it can be seen that the normal type 1 pattern of *Bt* israelensis is also found in a *Bt* canadensis T05a030 and in a *Bt* morrisoni, T08025.

In Figure 3.43, 4Btt, a strain of *Bt* tenebrionis is shown to have type 1 pattern in common with six strains of *Bt* morrisoni.

Figure 3.46 shows that an unusual strain of *Bt* darmstadiensis, T01246, has the same RAPD pattern as T18004, a strain of *Bt* kumamotoensis.

A strain of *Bacillus anthracis* (Ames – wild type) is shown in Figure 3.48, and the emetic toxin producing strain of *B. cereus* (F4810/72) has an almost identical RAPD pattern.

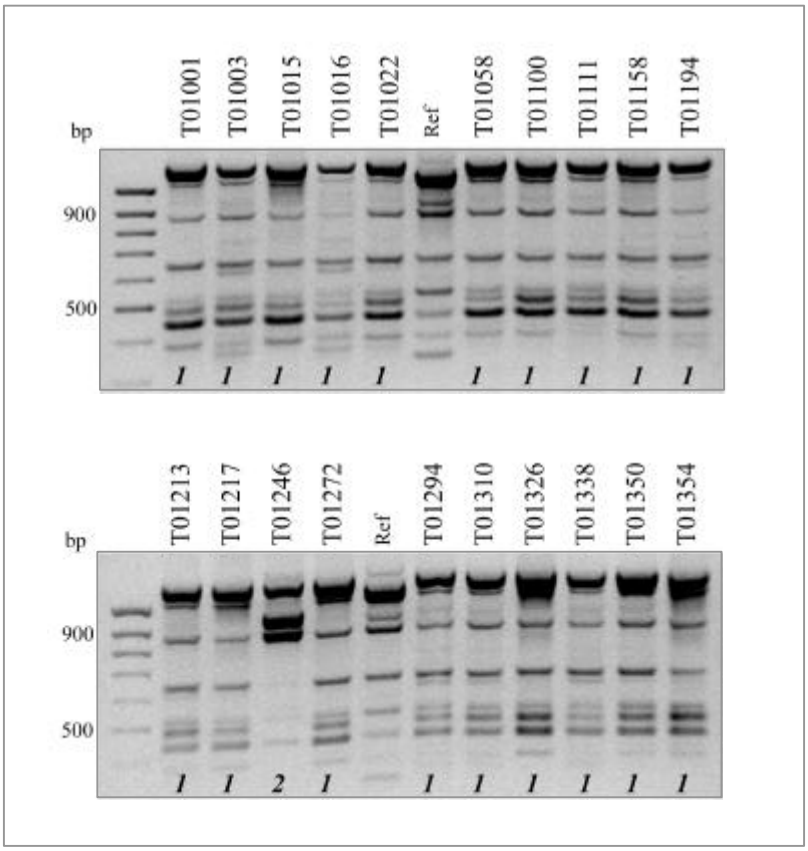


Figure 3.41 RAPD patterns for *Bt* thuringiensis.
The assigned pattern types are italicised

Bt thuringiensis

Strains of this serovar were very homogeneous, with only one strain giving a significantly different band pattern, out of 20 strains tested. This deviant strain was T01246, isolated from Iraq in 1984. Strains with identical RAPD patterns were isolated from many different countries, as shown below in Table 3.16.

Table 3.16 Countries of origin of *Bt* thuringiensis strains.

Strain	Country		Strain	Country		Strain	Country		Strain	Country
T01001	Canada		T01003	France		T01005	Bulgaria		T01016	England
T01022	USA		T01058	Switzerland		T01100	Finland		T01111	Finland
T01158	Czecho-slovakia		T01194	Japan		T01213	Czecho-slovakia		T01217	China
T01246	Iraq		T01272	China		T01294	Italy		T01310	Australia
T01326	Chile		T01338	Spain		T01350	Mexico		T01354	India

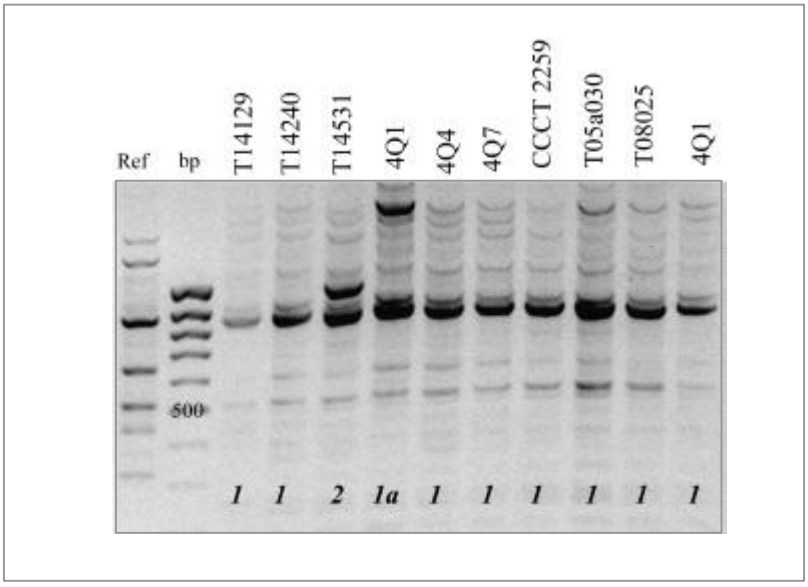


Figure 3.42 RAPD patterns for *Bt* israelensis and including two strains from different serovars - T05a030 *Bt* canadensis and T08025 *Bt* morrisoni. The assigned pattern types are italicised.

Bt israelensis

Bt israelensis has been studied extensively and was expected to be clonal (Priest *et al*., 1994). However, one strain, T1453 was found to have an additional band in the pattern. Subsequent MLST analysis found two different alleles in this strain. *Bt* israelensis CCCT 2259 had an MLST sequence type of 16; the typical ST for this serotype. Two strains from other serovars (T05a030, *Bt* canadensis from Iraq and T08025, *Bt* morrisoni from France) were allocated to ST-16 by MLST typing and were included in the RAPD gel. Their RAPD patterns were found to be identical to the RAPD patterns from *Bt* israelensis ST-16.

Table 3.17 Countries of origin of *Bt* israelensis strains.

strain	country		strain	country		strain	country		strain	country
T14129	Iraq		T14240	Colombia		T14531	Vietnam		4Q1	Israel
4Q4	Wild type BGSC		4Q7	Plasmid-cured		CCCT 2259	Brazil			

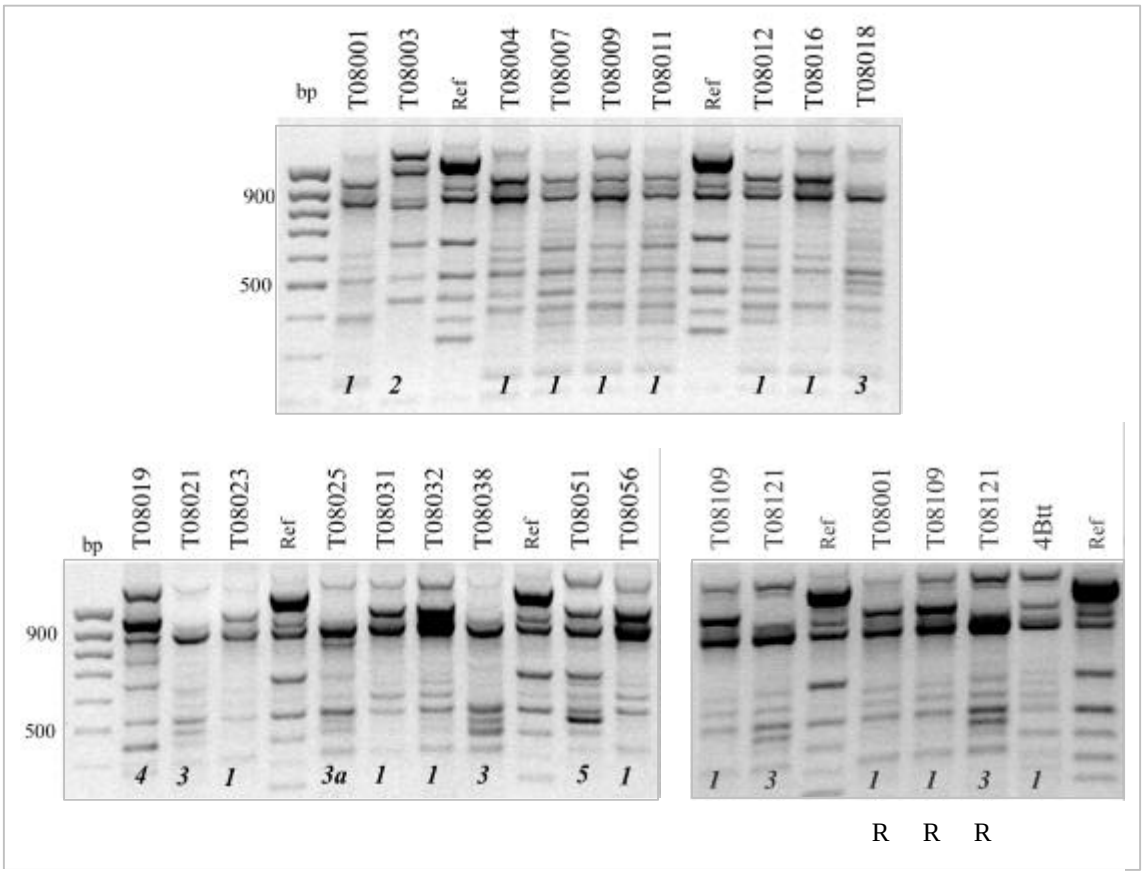


Figure 3.43 RAPD patterns for *Bt* morrisoni, and including strain 4Btt which is a *Bt* tenebrionis strain. The assigned pattern types are italicised

Bt morrisoni

Twenty strains of *Bt* morrisoni were examined, and a strain of *Bt* tenebrionis was included since it has the same serotype. A high degree of clonality was found in this serovar, with twelve of the strains belonging to band pattern type 1 (including *Bt* tenebrionis) and six in band pattern type 3. Band patterns 2, 4 and 5 had single representatives.

Table 3.18 Countries of origin of *Bt morrisoni* strains.

Strain	Country		Strain	Country		Strain	Country		Strain	Country
T08001	USA		T08003	Canada		T08004	USA		T08007	Czecho-slovakia
T08009	USA		T08011	France		T08012	Pakistan		T08016	Germany
T08018	Phillip-ines		T08019	Mexico		T08021	USA		T08023	Brazil
T08025	France		T08031	Brazil		T08032	Australia		T08038	Ghana
T08051	Canada		T08056	Algeria		T08109	Nigeria		T08121	Peru

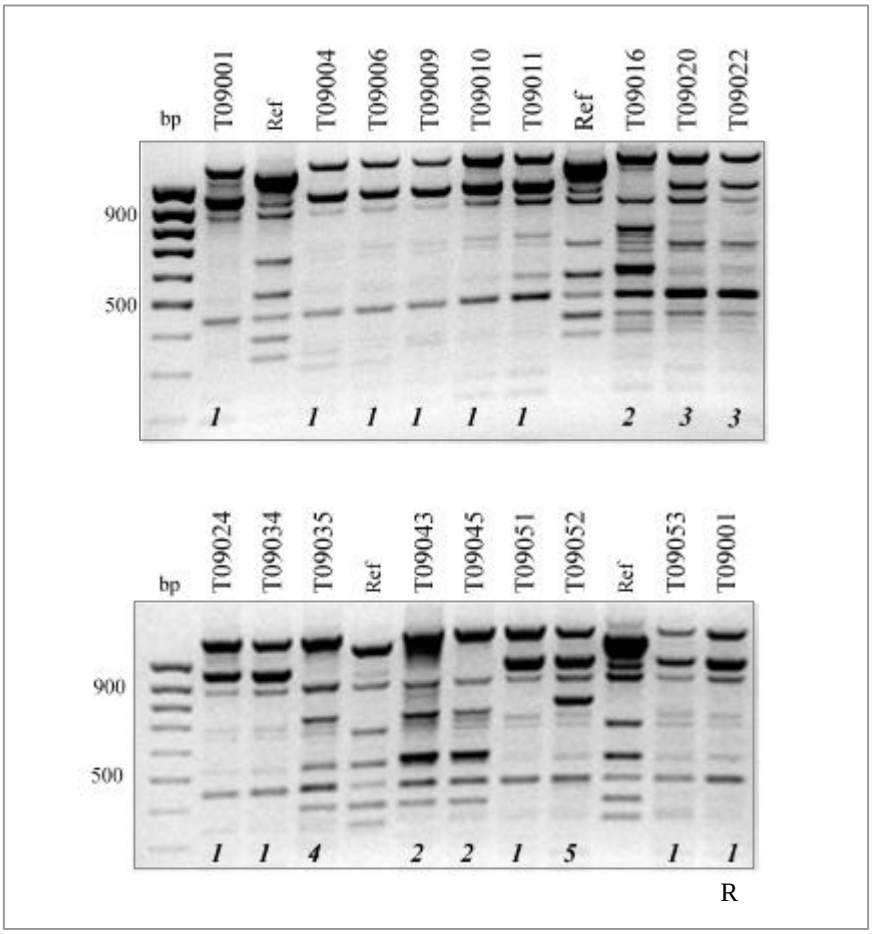


Figure 3.44 RAPD patterns for *Bt* tolworthi.
The assigned pattern types are italicised

Bt tolworthi

Of the 17 strains of *Bt* tolworthi tested, 10 strains were found to be type 1. There were three type 2 patterns and also two type 3 patterns. A relatively high degree of clonality was found in *Bt* tolworthi.

Table 3.19 Countries of origin of *Bt* tolworthi strains.

Strain	Country		Strain	Country		Strain	Country		Strain	Country
T09001	England		T09004	USA		T09006	England		T09009	England
T09010	USA		T09011	Iraq		T09016	Canada		T09020	Italy
T09022	Australia		T09024	Indonesia		T09034	Brazil		T09035	Chile
T09043	Thailand		T09045	Spain		T09051	Australia		T09052	Mexico
T09053	India									

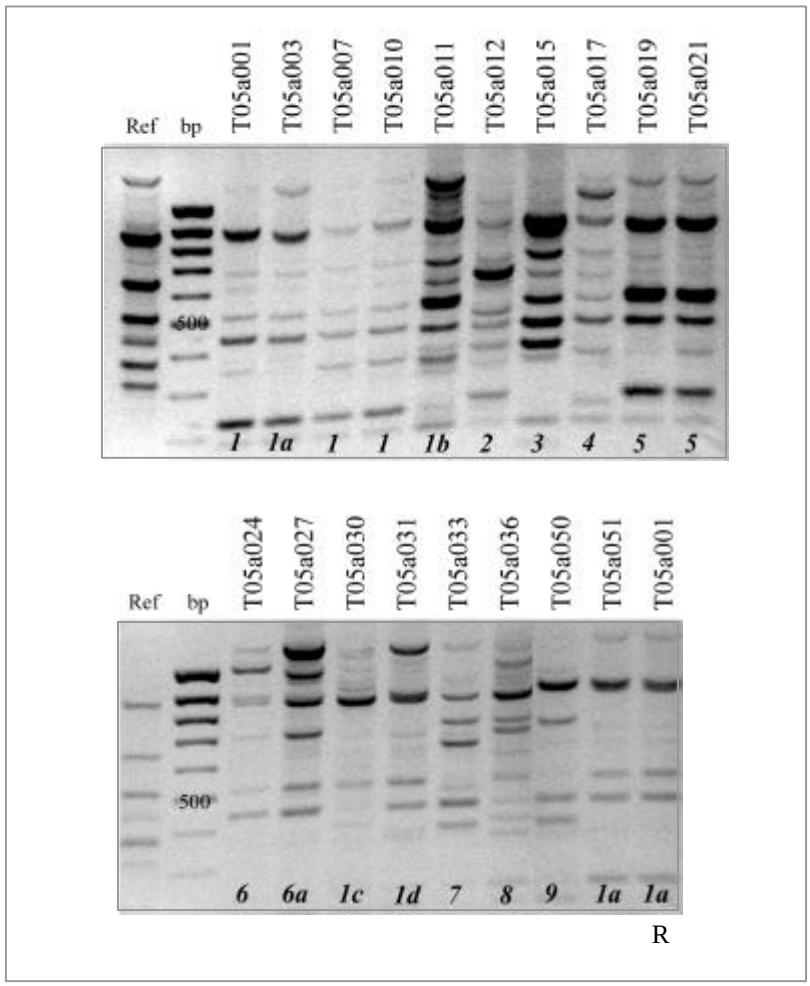


Figure 3.45 RAPD patterns for *Bt* canadensis
The assigned pattern types are italicised

Bt canadensis

Of the 18 strains of *Bt* canadensis assessed, eight were found to be type 1/1a/1b/1c/1d. The band patterns of the others were either singletons or occurred twice.

Table 3.20 Countries of origin of *Bt* canadensis strains

Strain	Country		Strain	Country		Strain	Country		Strain	Country
T05a001	Canada		T05a003	Chad		T05a007	Madagascar		T05a010	USA
T05a011	USA		T05a012	Czecho-slovakia		T05a015	USA		T05a017	Nigeria
T05a019	Pakistan		T05a021	France		T05a024	Czecho-slovakia		T05a027	Came-roon
T05a030	Iraq		T05a031	Mexico		T05a033	Kenya		T05a036	Korea
T05a050	Benin		T05a051	Colombia						

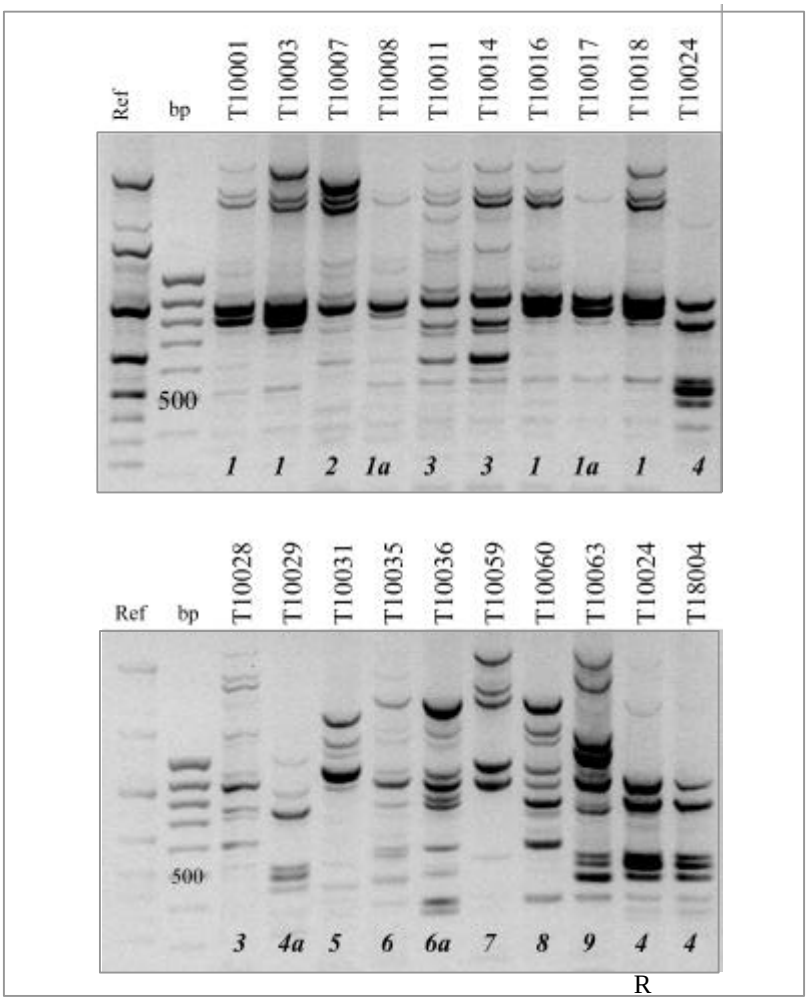


Figure 3.46 RAPD patterns for *Bt* darmstadiensis, and including T18004, an unusual strain of *Bt* kumamotoensis.
The assigned pattern types are italicised

Bt darmstadiensis

Eighteen strains of *Bt* darmstadiensis were examined. Six were found to have similar band patterns of type 1/1a. There were three type 3 patterns. An unusual strain of *Bt* kumamotoensis from Iraq (T18004), which has four MLST alleles in common with T10024 was included at the end of the RAPD gel. Both strains *Bt darmstadiensis* T10024 and *Bt kumamotoensis* T18004 gave RAPD pattern type 4. Another strain, T10029 had a similar pattern (4a).

Table 3.21 Countries of origin of *Bt* darmstadiensis strains.

Strain	Country		Strain	Country		Strain	Country		Strain	Country
T10001	Germany		T10003	Germany		T10007	USSR		T10008	Philippines
T10011	Japan		T10014	Finland		T10016	USA		T10018	Japan
T10024	Pakistan		T10029	Pakistan		T10031	Korea		T10035	Algeria
T10036	Spain		T10059	China		T10060	USA		T10063	Argentina
T10017	USA		T10028	Czech.						

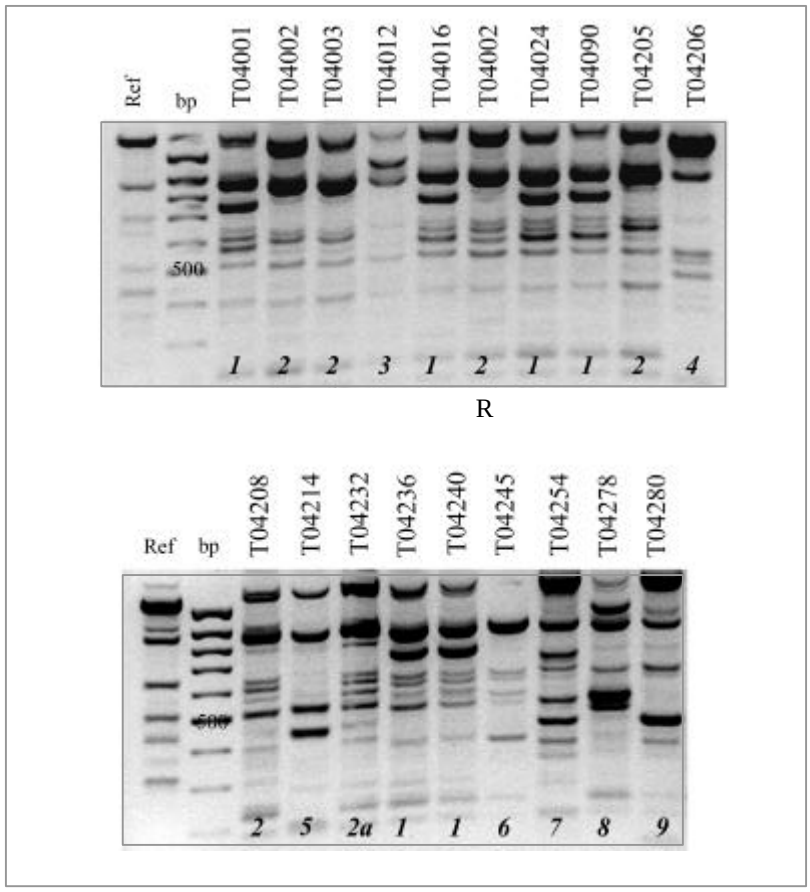


Figure 3.47 RAPD patterns for *Bt* sotto.
The assigned pattern types are italicised.

Bt sotto

Of the 18 strains of *Bt* sotto were examined, six were band type 1, six were band type 2/2a and there were seven individual types. Some clonality was evident, but there was diversity too.

Table 3.22 Countries of origin of *Bt* sotto strains

Strain	Country		Strain	Country		Strain	Country		Strain	Country
T04001	Canada		T04002	Canada		T04003	France		T04012	USA
T04016	Pakistan		T04024	Pakistan		T04090	Pakistan		T04205	USA
T04206	Checho-slovakia		T04208	Japan		T04214	Mexico		T04232	Korea
T04236	Indonesia		T04240	Indonesia		T04245	Kenya		T04254	Spain
T04278	Pakistan		T04280	Thailand						

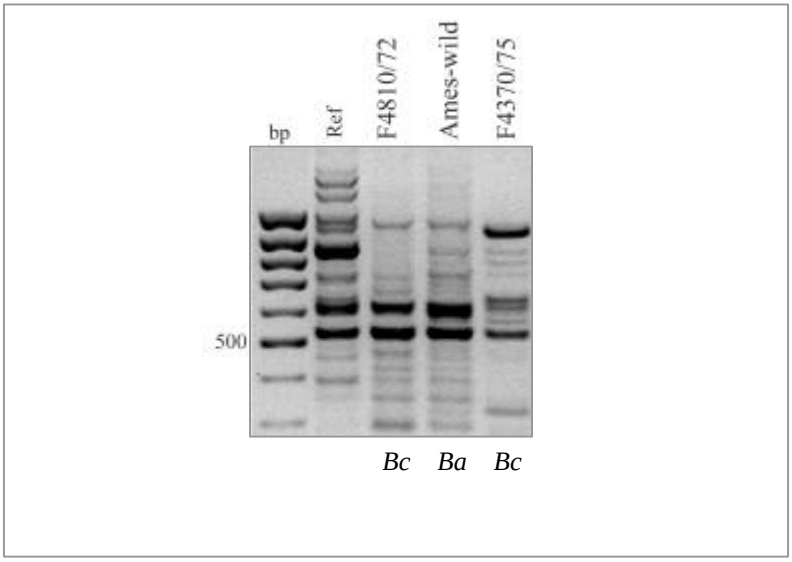


Figure 3.48 RAPD patterns for *B. cereus* and *B. anthracis (*Ames – wild type).

B. anthracis and two *B. cereus* strains

RAPD products were electrophoresed with DNA from *B. anthracis* (Ames – wild type), an emetic toxin producing strain of *B. cereus* (F4810/72), and a *B. cereus* which was the closest strain to *B. anthracis* in the MLST dataset (F4370/75). This gel showed that the emetic toxin producing strain of *B. cereus* had a very similar band pattern to *B. anthracis*, but strain F4370/75 was very different.

F4370/75 was positioned in lineage Cereus III, (close to *B. anthracis*) on the phylogenetic tree shown in Figure 3.9, but the emetic toxin producing strains were further away, in Cereus II.

In summary, these data show that while serotyping provides a reasonable assessment of genomic similarity for some clonal serovars such as *Bt* thuringiensis, there is however, relatively little correlation between serovar and RAPD pattern for most of these groups**.**

3.3 Detection of crystal toxin genes by PCR.

The crystal toxins of *B. thuringiensis* are active against particular types of insects. Cry1 is active against *Lepidoptera* (butterflies and moths) larvae. Cry3 targets *Coleoptera* insects (beetles) and Cry4 is active against *Diptera* insects (mosquitos and blackflies). In this section, various strains of *B. thuringiensis* were screened for particular c*ry* genes by PCR to determine if there was any correlation between RAPD type, serovar and *cry* gene content

B. thuringiensis serovars thuringiensis and tolworthi were tested for the presence of crystal toxin genes c*ry1, cry1b, cry1e* and *cry3a*. and *B. thuringiensis* serovar morrisoni strains were tested for the presence of *cry1, cry3a* and c*ry4* genes.

A *Bt* israelensis strain and a *Bt* tenebrionis were also examined for *cry1, cry3a* and *cry4* genes.

Table 3.23 Primers for *B. thuringiensis* crystal toxin genes.

Target	Forward	Reverse	Expected product size(bp)
cry1	gra/cry/d	gra/cry/r	558
cry1b	CJ8	CJ9	367
cry1e	1E	1 (-)	1137
cry3a	Un3 (d)	EE-3A(r)	951
cry4	Dip1A	Dip1B	797

Full details of the *cry* gene primers are given in ‘Materials and Methods section 2.7’.

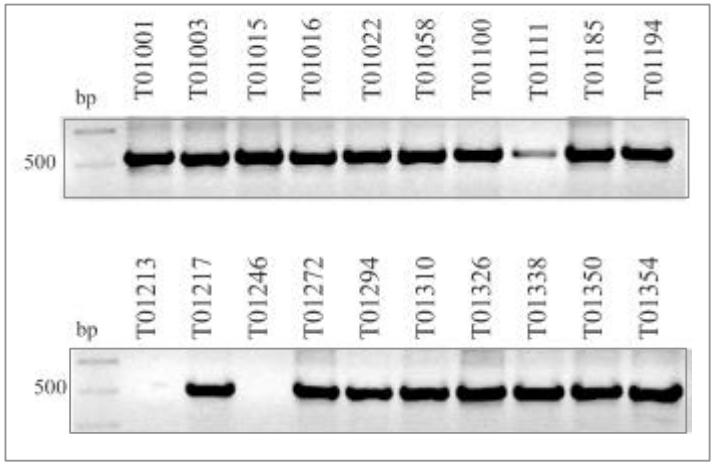


Figure 3.49 Amplification of *cry1* genes from DNA of *Bt* thuringiensis strains. (Primers were gra/cry/d and gra/cry/r , amplified product is 558 bp)

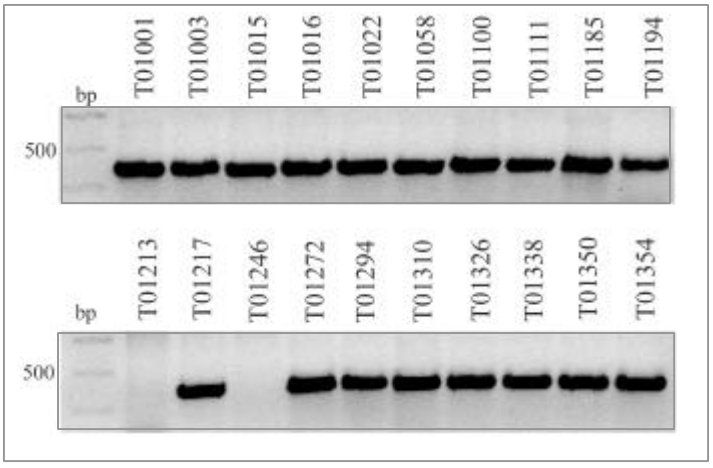


Figure 3.50 Amplification of c*ry1b* genes from DNA of *Bt* thuringiensis strains. (Primers are CJ8 and CJ9, amplified product is 367 bp)

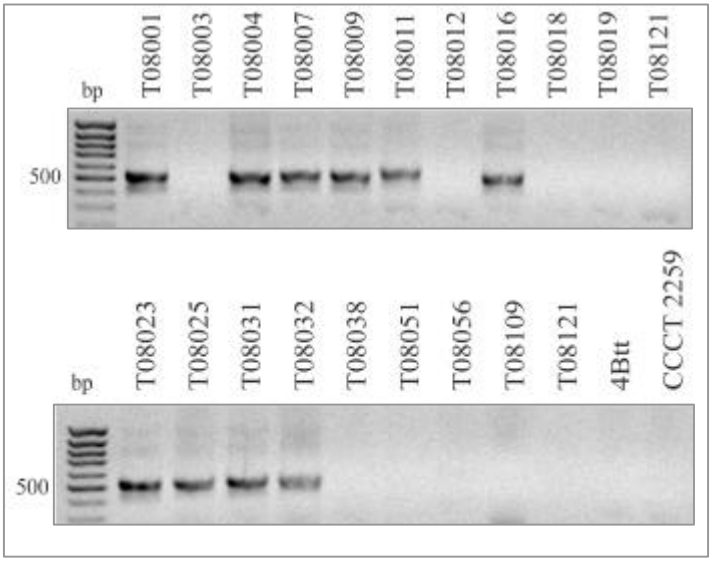


Figure 3.51 Amplification of c*ry1* genes from DNA of *Bt* morrisoni strains. (Primers were gra/cry/d and gra/cry/r, amplified product is 558 bp)

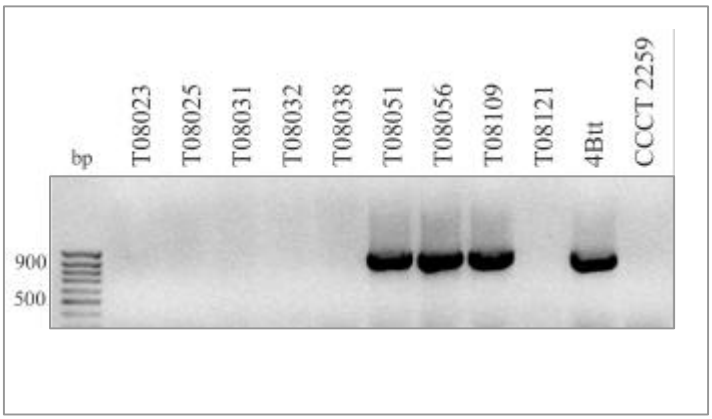


Figure 3.52 Amplification of *cry3a* genes from DNA of *Bt* morrisoni, *Bt* tenebrionis (4Btt) and *Bt* israelensis (CCCT 2259).
(Primers were Dip Un 3 (d) and EE-3Aa(r), amplified product is 951 bp)

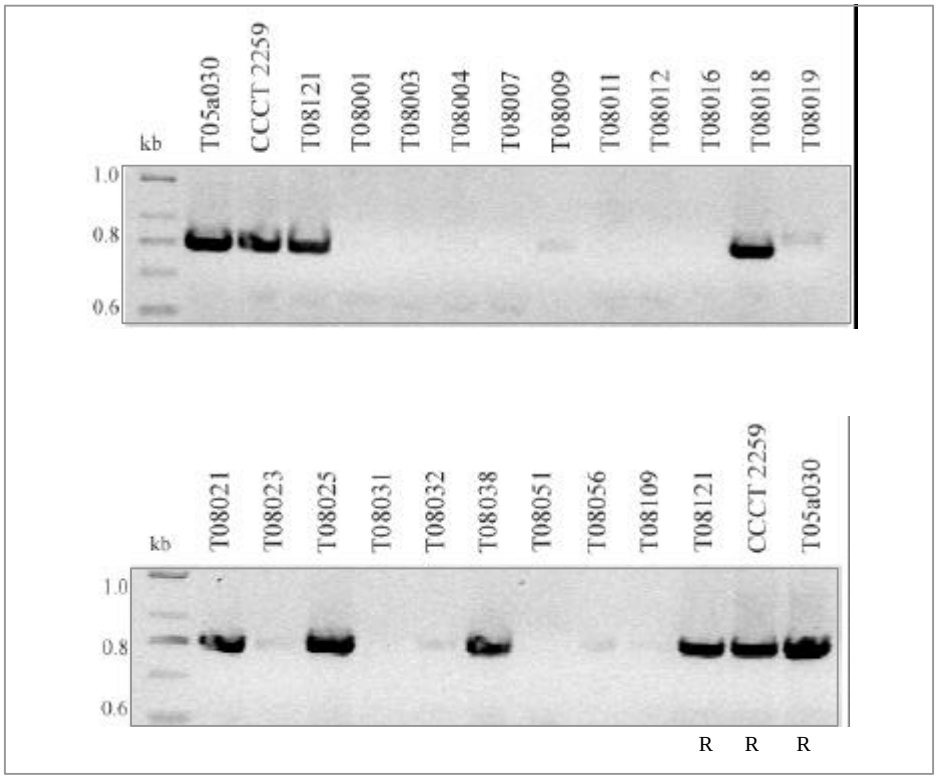


Figure 3.53 Amplification of *cry4* genes from DNA of *Bt* morrisoni , *Bt* canadensis (T05a030) and *Bt* israelensis (CCCT 2259)
Primers were Dip1A and Dip1B, amplified product is 797 bp.

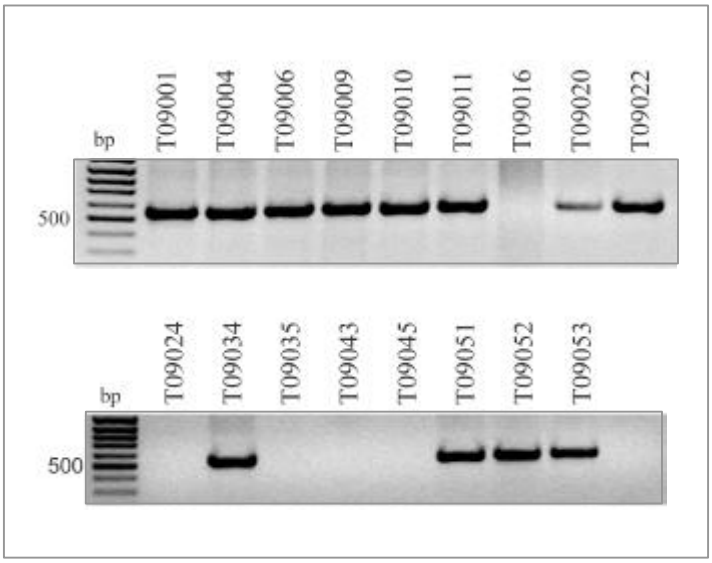


Figure 3.54 Amplification of *cry1* genes from DNA of *Bt* tolworthi strains. (Primers were gra/cry/d and gra/cry/r, amplified product is 558 bp)

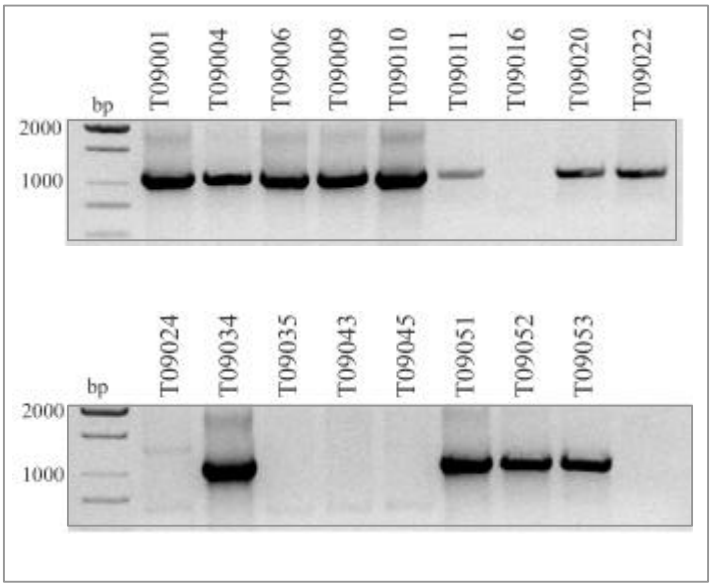


Figure 3.55 Amplification of *cry1e* genes from DNA of *Bt* tolworthi. (Primers were 1E and 1(-), amplified product is 1137 bp)

3.4 Combined results for MLST, RAPD and *cry* genes

Table 3.24 *Bt* thuringiensis. Combined results of crystal toxin genes, RAPD patterns, MLST and alleles

Bt thuringiensis	* Crystal toxin genes	RAPD pattern	MLST sequence type
T01001	1, 1b	1	10
T01003	1, 1b	1	
T01015	1, 1b	1	10
T01016	1, 1b	1	
T01022	1, 1b	1	10
T01058	1, 1b, 3a	1	10
T01100	1, 1b	1	
T01111	1, 1b	1	
T01185	1, 1b	1	
T01194	1, 1b	1	
T01213	Not amplified	1	
T01217	1, 1b	1	
T01272	1, 1b	1	
T01294	1, 1b	1	
T01310	1, 1b	1	
T01326	1, 1b	1	10
T01338	1, 1b	1	
T01350	1, 1b	1	
T01354	1, 1b	1	
T01246	Not Amplified	2	48

All *Bt* thuringiensis strains were examined for the presence of *cry1*, *cry1b*, *cry1e* and *cry3a* genes. The results are ordered according to RAPD pattern.

B. thuringiensis serovar thuringiensis is the most clonal of the serovars examined, with only one unusual strain revealed by RAPD (T01246) which also had a unique MLST sequence type. Two strains, T01246 and T01213 did not amplify the *cry1* genes, of which one (T01246, ST-48) was the atypical strain. All 7 alleles from T01246 were different from those found in ST-10.

An interesting discovery was that strain T01058, with RAPD pattern 1 and the most prevalent MLST sequence type of 10 was found to have the *cry3a* gene, which is specific to Coleopteran insects. This strain also had the *cry1* genes and therefore could be toxic to insects from both Lepidopteran and Coleopteran orders.

Table 3.25 *Bt* morrisoni (and other ST-16 strains). Combined results of crystal toxin genes, RAPD patterns, MLST and alleles

Bt morrisoni	* Crystal toxin genes	RAPD pattern	MLST sequence type
T08001	1	1	23
T08004	1	1	
T08007	1	1	
T08009	1	1	23
T08011	1	1	
T08012	not amplified	1	23
T08016	1	1	69
T08023	1	1	23
T08031	1	1	23
T08032	1	1	
T08056	3a	1	
T08109	3a	1	
T08003	not amplified	2	64
T08018	4	3	
T08021	4	3	
T08038	4	3	
T08121	4	3	67
T08025	1 4	3a	16
T08019	not amplified	4	63
T08051	3a	5	70
Bt tenebrionis			
4 Btt	3a	1	23
Bt israelensis			
** CCCT 2259	4	N/A	16
Bt canadensis			
** T05a030	4	N/A	16

* All *Bt* morrisoni strains were examined for the presence of *cry1, cry 3a* and *cry4*. Results are ordered according to RAPD pattern.

** *Bt israelensis* CCCT 2259 and *Bt* canadensis T05a030 were tested for the presence of *cry4* genes as they were assigned ST-16. These two strains were not run on the same RAPD gel as the *Bt* morrisoni strains (Fig 3.43) and their RAPD patterns are not comparable.

The *Bt* morrisoni serovar was highly clonal, with most strains producing RAPD pattern 1.

DNA from three strains amplified the Coleopteran specific *cry3a* gene.

Five strains which were RAPD pattern 3/3a, were found to have the Dipteran specific *cry4* gene, suggesting that these strains are actually slight variants of *Bt* israelensis with the same *cry4* genes.

The DNA from one strain, T08025, amplified *cry4* and had the same MLST sequence type (16) as *Bt* israelensis.

DNA from *Bt* tenebrionis amplified the *Cry3a* gene, but had RAPD pattern 1, in common with the majority of the *cry1* containing *Bt* morrisoni serovars.

Allele *glp*-15 was found in all strains examined, except strain T08003. The *ilvD*-7 allele was present in the majority of the *Bt* morrisoni strains, but T08003, T08016 and T08051 had allele *ilvD*-36, suggesting possible horizontal gene transfer involving *ilvD*.

Table 3.26 *Bt* tolworthi. Combined results of crystal toxin genes, RAPD patterns, MLST and alleles.

Bt tolworthi	* Crystal toxin genes	RAPD pattern	MLST sequence type
T09001	1, 1e	1	
T09004	1, 1e	1	
T09006	1, 1e	1	
T09009	1, 1e	1	
T09010	1, 1e	1	22
T09011	1, 1e	1	22
T09024	Not amplified	1	22
T09034	1, 1e	1	
T09051	1, 1e	1	
T09053	1, 1e	1	
T09016	Not amplified	2	68
T09043	Not amplified	2	68
T09045	Not amplified	2	
T09020	1, 1e	3	
T09022	1, 1e	3	22
T09035	Not amplified	4	
T09052	1, 1e	5	22

All *Bt* tolworthi strains were examined for the presence of *cry1, cry1b, cry1e* and *cry3a*. Results are ordered according to RAPD pattern.

All strains with RAPD pattern type 1 had an MLST sequence type 22 where examined. However, two strains, T09022 and T09052 had the same ST-22, but their RAPD patterns were slightly different (type 3 and type 5 respectively). DNA from three strains with RAPD pattern 2 did not amplify the *cry1* gene, and two of these strains (T09016 and T09043), which have been MLST typed were both ST-68.

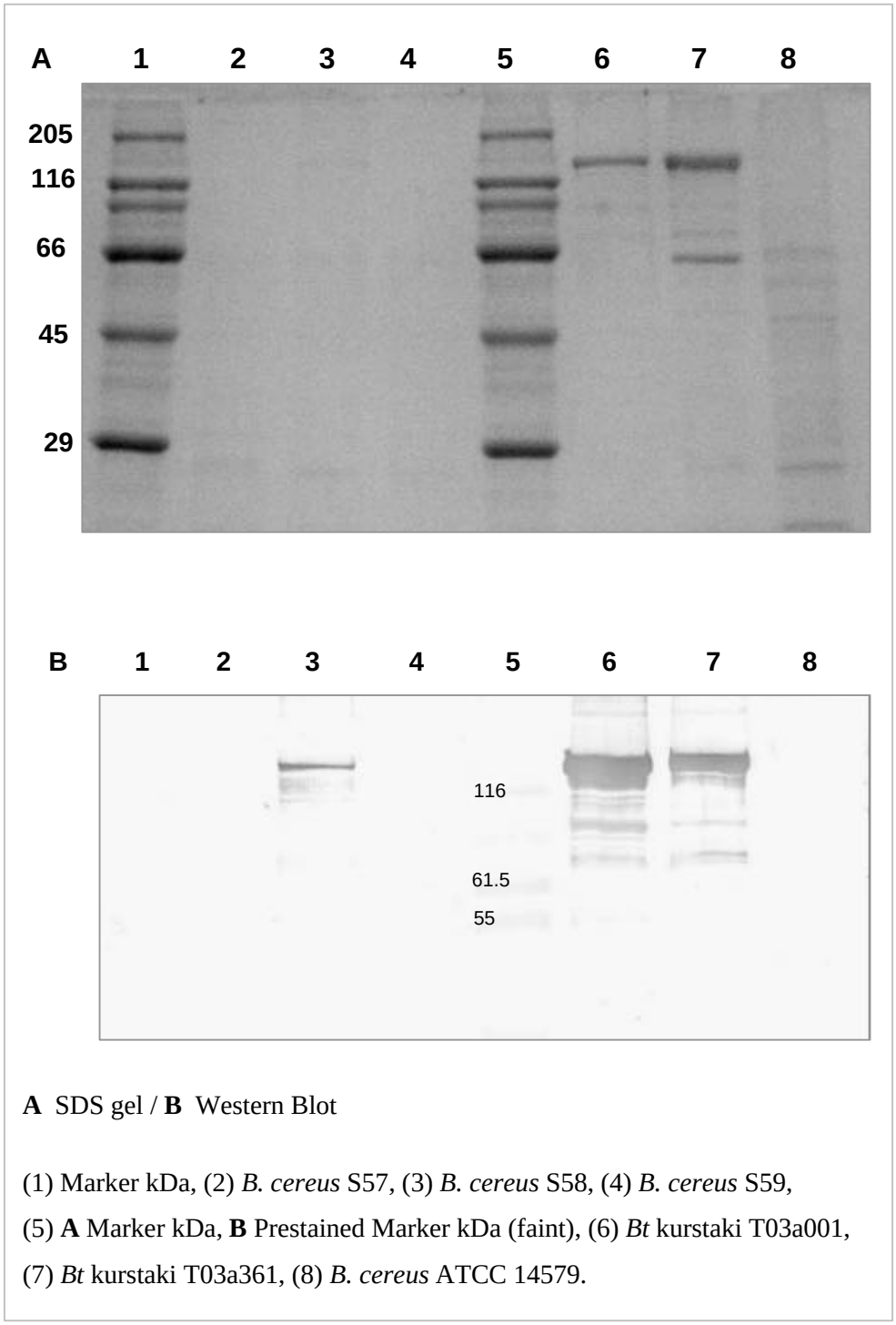
In summary, there was generally a correlation between RAPD pattern type, MLST sequence type and the presence of crystal toxin genes, but there were exceptions. Nevertheless, MLST / RAPD provided more robust indicators of *cry* gene content than serotyping.

3.5 SDS-PAGE and Western Blot

Several strains of *B. thuringiensis* kurstaki had identical RAPD patterns and MLST sequence types to three strains of *B. cereus* (S57, S58, S59) - all from Maryland. *B. thuringiensis* differs from *B. cereus* in that it produces a crystal protein which is toxic to insects. The *B. cereus* strains were examined for crystal toxin synthesis.

The strains were grown until sporulation, and crystals prepared according to the method given by Lecadet *et al.* (1992). The SDS gel is shown in Figure 3.56, and proteins of the expected size of Cry1 (120 kD) were isolated from the *Bt* kurstaki strain, used as a control. Similar proteins were not evident in the *B. cereus* strains.

A Western Blot was carried out to determine if the protein was recognised by Cry1 antibody. The protein from the reference strain reacted with the antibody, but unfortunately the antibody also reacted with some other proteins present in the gel. One of the *B. cereus* strains, S58, also reacted with the Cry1 antibody (lane 3).



A SDS gel / **B** Western Blot

(1) Marker kDa, (2) *B. cereus* S57, (3) *B. cereus* S58, (4) *B. cereus* S59,
(5) **A** Marker kDa, **B** Prestained Marker kDa (faint), (6) *Bt* kurstaki T03a001,
(7) *Bt* kurstaki T03a361, (8) *B. cereus* ATCC 14579.

Figure 3.56 (A) SDS - PAGE and (B) Western Blot of *Bacillus thuringiensis* kurstaki and *Bacillus cereus* strains with the same ST.

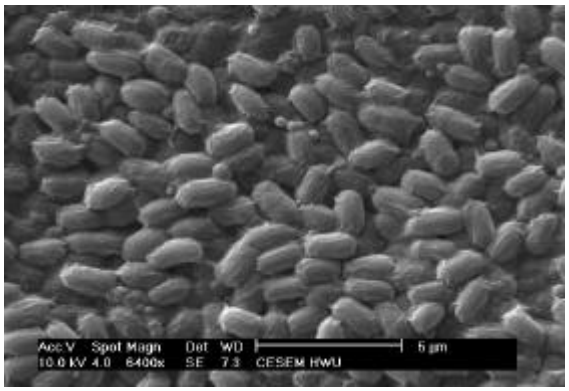
3.6 ELECTRON MICROSCOPY

To resolve the uncertainty about whether the *B. cereus* strains produced crystal proteins or not, a spore/crystal extract was viewed under the scanning electron microscope, and is shown in Figure 3.57.

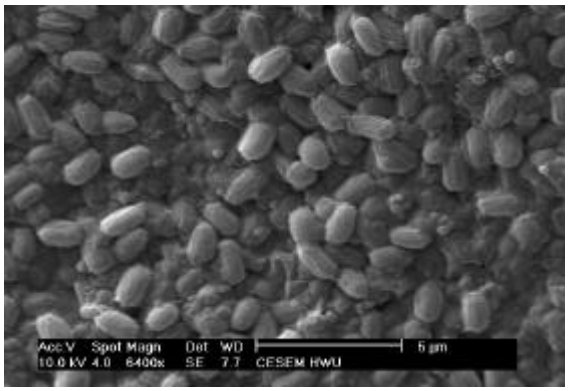
The *Bt* kurstaki strain had abundant bipyramidal crystals, but no crystals were seen in the *B. cereus* preparations.

The three *B. cereus* strains have been kept in the laboratory for a long time, and reports that *B. thuringiensis* strains can loose the ability to produce crystal proteins over time have been published (Gordon *et al.*,1973). It may be that the strains collected as *B. cereus* a long time ago, were actually *B. thuringiensis* and have now lost their ability to make crystal protein, as existing in a laboratory environment does not require the ability to be parasitic to insect larvae.

Electron microscope photographs of *Bacillus cereus* and *Bacillus thuringiensis* strains, with identical MLST sequence types, and RAPD patterns.

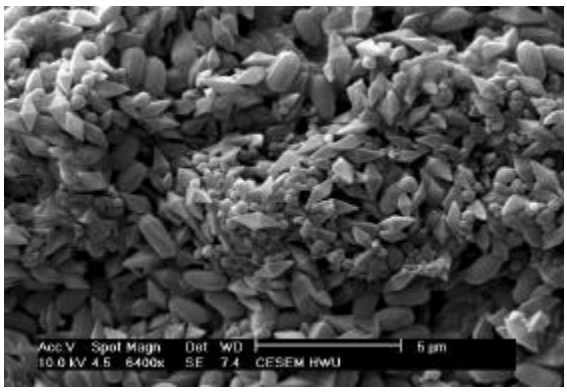


Bacillus cereus strain S57.
Abundant spores, but no evidence of crystals.

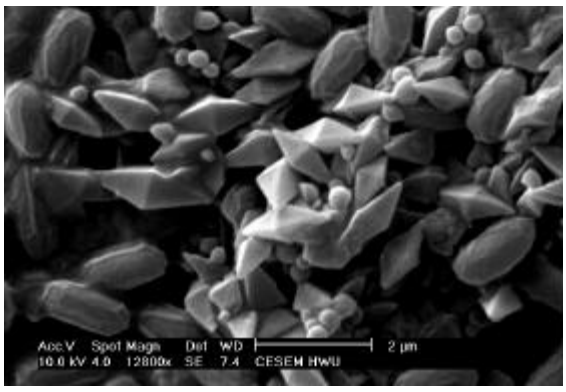


Bacillus cereus strain S58.

The nature of the small spherical bodies (approx.0.04 µm in diameter), present in both *B. cereus* and *B. thuringiensis* preparations, is unknown, but they may be aggregates of carbohydrate or proteins.



Bacillus thuringiensis, strain T03a001.
Spores and bipyramidal crystals



Bacillus thuringiensis, strain T03a001.
Spores and bipyramidal crystals.
(Larger magnification)

Figure 3.57 Electron microscope photographs of spore/crystal protein extracts from *B.thuringiensis* kurstaki and *B.cereus* S57 and S58.

3.7 Semi-quantitative assay of toxicity of *Bt* strains to mosquito larvae

Several strains of *B. thuringienis* were assessed for mosquitocidal activity. A strain of *Bt* israelensis, CCCT 2259 with the usual RAPD pattern (1_isr) and ST-16, was compared with a strain of *Bt* israelensis with an unusual RAPD pattern (2_isr) and ST-65. Three atypical strains from other serovars were included as controls with no known insect toxicity. (No PCR product had been amplified when they were examined using primers for *cry1, cry3 and cry4* genes): -

Table 3.27 Serial dilutions for mosquitocidal assay

cell	1	2	3	4	5
dilution	1 ml culture + 1ml water	1 ml from cell 1 + 1 ml H_20	1 ml from cell 2 + 1 ml H_20	1 ml from cell 3 + 1 ml H_20	1 ml from cell 4 + 1 ml H_20

Tests were carried out in duplicate in plastic trays 10 cm^2, divided into 25 cells, each 2 cm^3 (section 2.12). Five healthy, active mosquito larvae were placed in each cell, and observed at intervals. After one hour, all five larvae in cell 1, with *Bt* israelensis CCCT 2259, and *Bt* israelensis T14531were dead. There was some slight activity in cell 5 (tails twitching). After one hour and forty-five minutes, all five larvae in cells 1 to 5, containing *Bt* israelensis strains CCCT 2259 and T14531 were dead even at the lowest dilution. All larvae in cells containing *Bt* strains other than *Bt* israelensis were swimming actively, and were still unaffected after 24 hours, as shown in Table 3.28.

Table 3.28 Effect of *Bt* israelensis on mosquito larvae.

Strain	Species/serovar	ST	*cry* genes	mosquitocidal activity
T01246	*Bt* thuringiensis	48	not detected	-
T10024	*Bt* darmstadiensis	57	not detected	-
CCCT 2259	*Bt* israelensis	16	4	+
T14531	*Bt* israelensis	65	4	+
T18004	*Bt* kumamotoensis	60	not detected	-

This experiment showed that both *Bt* israelensis T14531, and *Bt* israelensis CCCT 2259, which differ by two alleles, have equivalent mosquitocidal activity. The unusual strains of *Bt* thuringiensis, *Bt* darmstadiensis and *Bt* kumamotoensis did not have any effect on mosquito larvae.

3.8 Starch Hydrolysis

A selection of strains from the collection was tested for their ability to hydrolyze starch.

B. cereus strains are generally considered to be positive for starch hydrolysis. However, in this study, 108 strains were tested and 38 (35%) were found to be negative. It should be mentioned that some of the results were unclear, as there appeared to be a spectrum of results, varying from strongly positive, through moderately and weakly positive to negative.

An inability to hydrolyse starch has been reported (Agata *et al.*, 1996) to be a characteristic feature of the emetic toxin producing strains of *B. cereus* and this was confirmed, but many other strains, including some *B. thuringiensis* strains also gave negative results for starch hydrolysis.

The complete results of starch hydrolysis and the presence of enterotoxin genes is given in Table 3.29.

3.9 Enterotoxin genes

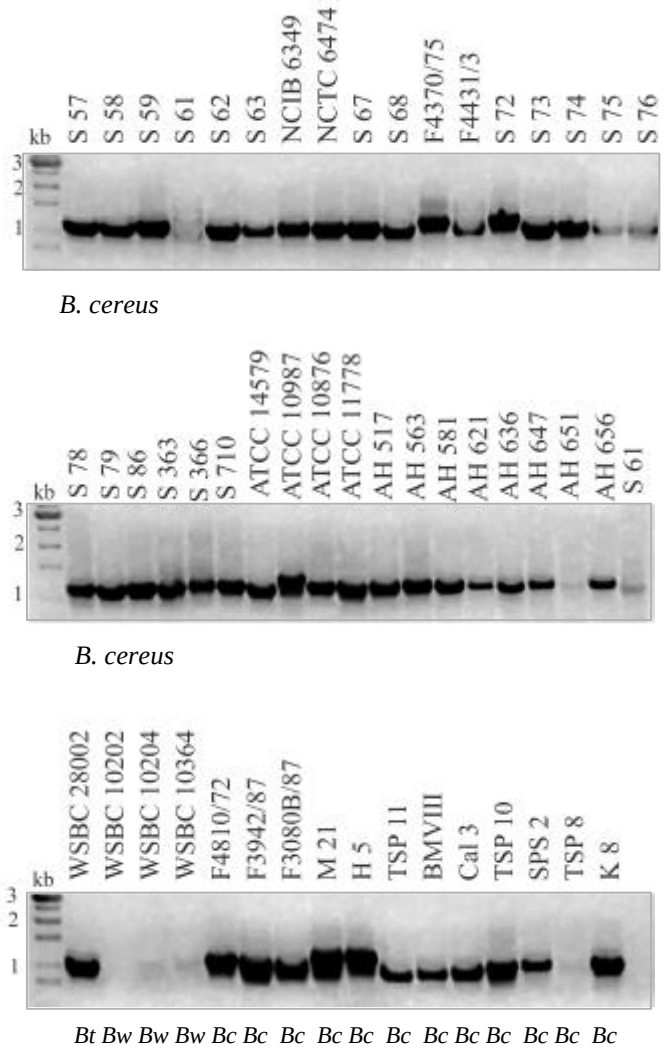
Bacteria from the *B. cereus* group have been known to cause two types of food poisoning. The emetic type of toxin has been found to be a cyclic peptide, cereulide (Agata *et al.,* 1995). Recent work has suggested that cereulide is produced by an enzyme complex known as non-ribosomal peptide synthetase (NRPS) (Horwood *et al.,* 2004). Emetic producing strains of *B. cere*us form a distinct clonal group (Pirrtijarvi *et al.,* 1999; Priest *et al.,* 2004; Ehling-Schulz *et al.,* 2005). The diarrhoeal form of food poisoning is caused by two gene complexes or operons; a non-haemolytic complex (*nhe*) and a haemolytic complex (*hbl*). *nhe* has been found in nearly all *B. cereus* strains, but *hbl* is present in about half of the *B. cereus* strains examined (Lindbeck *et al.,* 2004). In 2000, Gaviria *et al.* found that all of the 74 *B. thuringiensis* strains examined by PCR had the *nhe* gene and 88% had *hbl*, suggesting that these toxin genes may be more common in strains classified as *B. thuringiensis* than in *B. cereus.*

A wide range of strains from the *B. cereus* group of bacteria were tested for the presence of the enterotoxin gene complexes *nhe* and *hbl*, by PCR, using published primers (Granum *et al.,* 1999. Økstad *et al.,* 1999). Details of all primers are described in the 'Materials and Methods' section 2.13. New primers were designed to try to amplify the three genes which make up the *nhe* operon, and to amplify three of the four genes from the *hbl* operon.

A selection of gels showing the results of the PCR amplification for the enterotoxin genes is shown in Figures 3.58 to 3.61. Table 3.29 lists all the results for the enterotoxin gene PCRs, and includes the results of the starch hydrolysis test. (Table 3.29 is in order of MLST sequence type).

In this study, 89% of the 128 strains tested were positive for *nhe* (*BC* genes) and 75% were positive for *hbl* (*CD* genes).

Figure 3.62 shows the distribution of the enterotoxin genes on the NJ phylogenetic tree of 88 STs. The presence of both *hbl* and *nhe* genes is more common in clade 2, which is predominantly composed of *B. thuringiensis* strains.



There is no PCR product (or only a faint one) with the *B. weihenstephanensis* strains

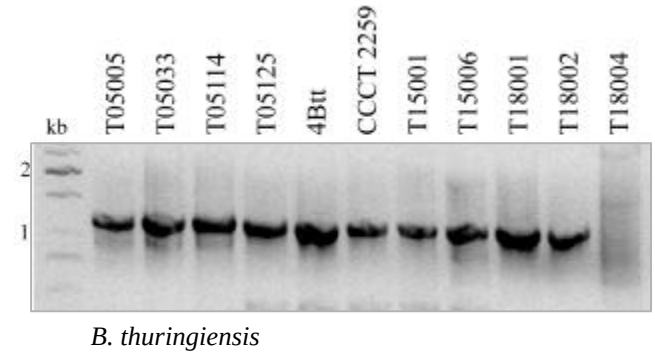


Figure 3.58 Examples of gels showing PCR amplification of *nhe* gene (*BC* component) Primers are nheA1F and nhe B1R, amplified product is 1200 bp.

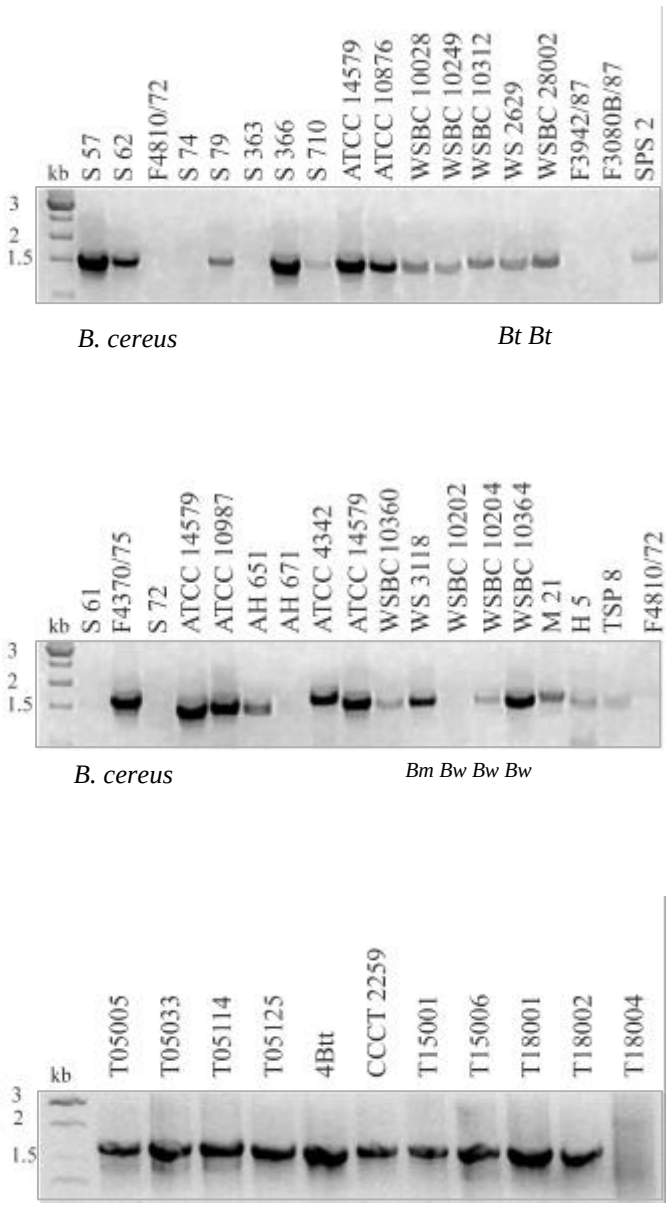


Figure 3.59 Examples of gels showing PCR amplification of *nhe* gene (*ABC* component). (Primers are nheA1301F and nhe C3017R, expected amplified product is 11716 bp)

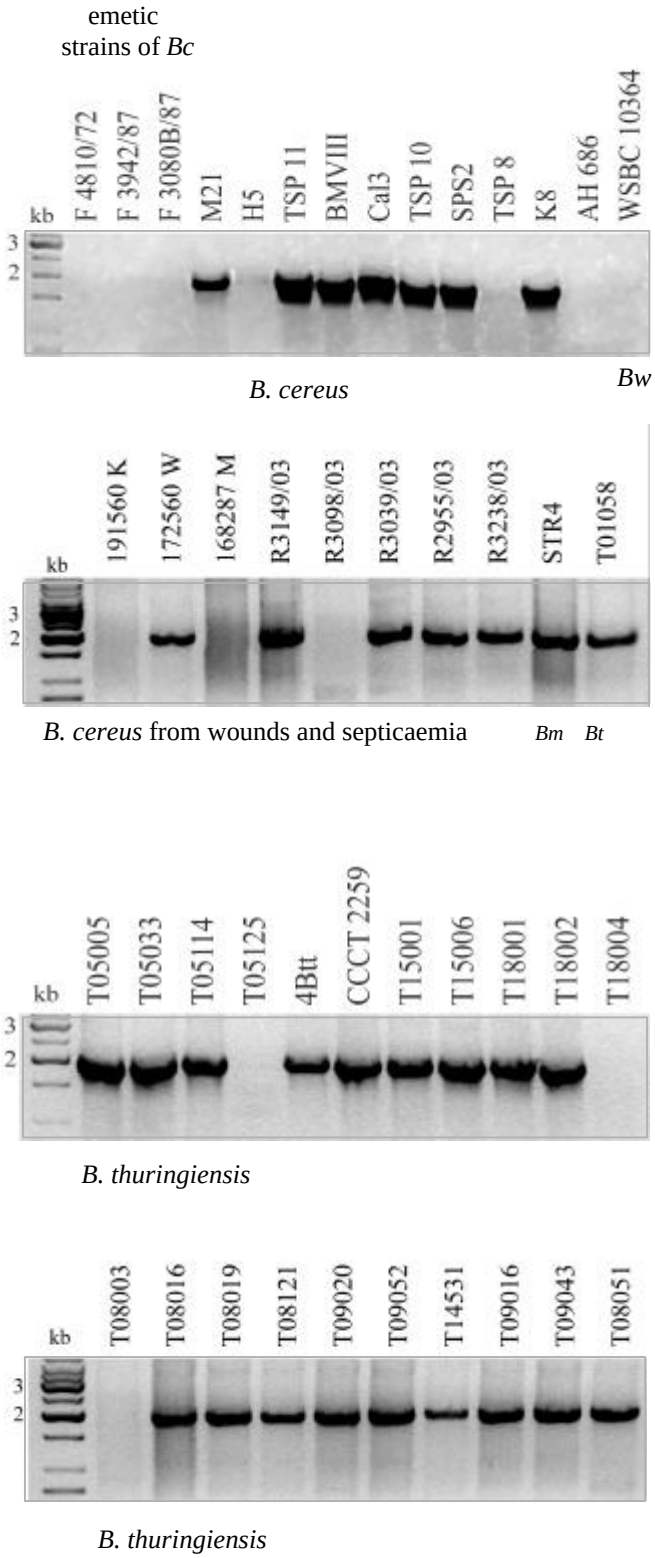


Figure 3.60 A selection of gels showing PCR amplification of *hbl* gene (*CD* component). Primers are hbl CF1 and hbl DR1; amplified product is 1900 bp.

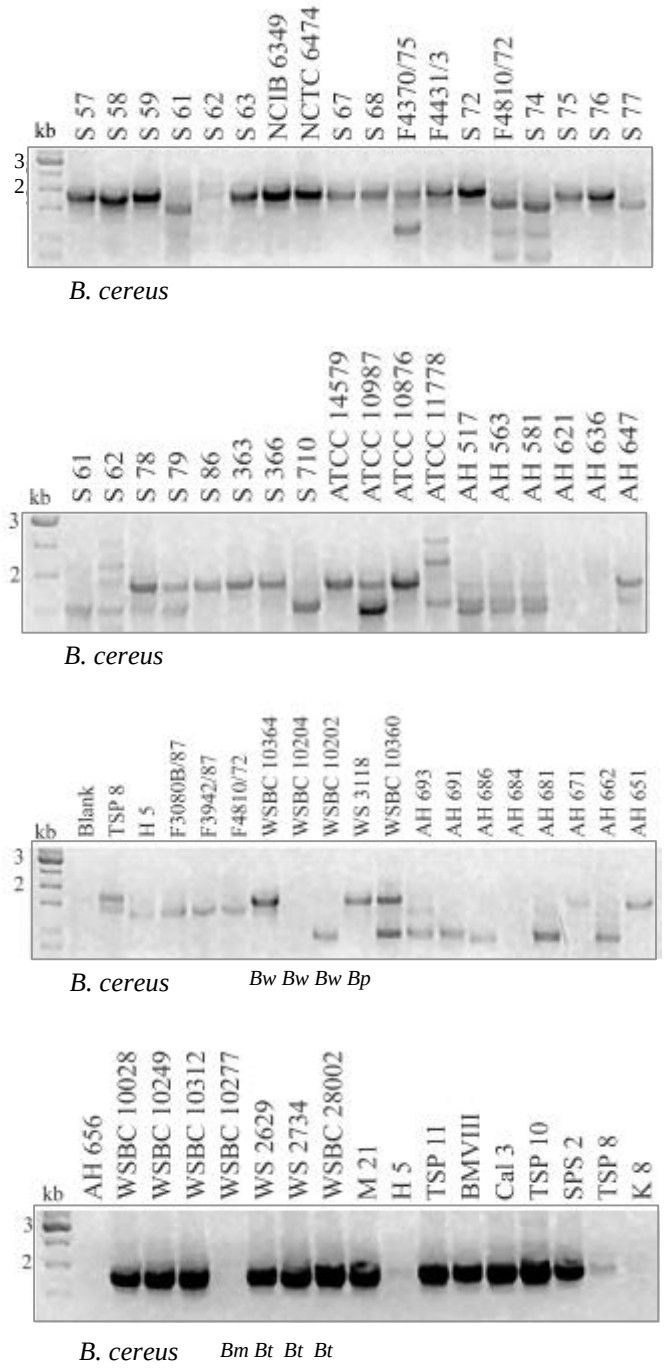


Figure 3.61 Examples of gels showoing PCR amplification of hbl gene (*CDA* component) Primers are hbl C11729F and hbl A1362R; expected amplified product is 1892 bp.

Table 3.29 Presence of enterotoxin genes in strains of *B. cereus, B. thuringiensis, B. mycoides, B. pseudomycoides* and *B. weihenstephanensis* (in order of Sequence Type).

strain	**species / serovar**	**MLST ST**	**Starch hydrolysis**	**nhe *B C* genes**	**nhe *A B C* genes**	**hbl *C D* genes**	**hbl *C D A* genes**
		Predicted Band size (bp)		*1200*	*1635*	*1900*	*1892*
*Ames	*anthracis*	**1**		+		-	
ATCC14579	*cereus*	**4**	+	+	+	+	+
m1545	*cereus*	**5**	-	+	-	-	-
m1564	*cereus*	**6**	-	+	-	-	+
M21	*cereus*	**7**	-	1300	+	+	+
S57	*cereus*	**8**	+	+	+	+	+
S58	*cereus*	**8**	+	+	+	+	+
S59	*cereus*	**8**	+	+		+	+
T03a001	kurstaki	**8**	+	+		-	
T03a075	kurstaki	**8**	+	+		+	
T03a172	kurstaki	**8**	+	+		-	
T03a287	kurstaki	**8**	+	+		+	
T03a361	kurstaki	**8**	+	+		-	
NCTC 6474	*cereus*	**9**	+	+	-	+	+
T01001	thuringiensis	**10**	+	+		+	
T01015	thuringiensis	**10**	-	+		-	
T01022	thuringiensis	**10**	+	+		+	
T01058	thuringiensis	**10**		-		+	
T01326	thuringiensis	**10**	+	+		+	
T04002	sotto	**12**	+	+		+	
T04016	sotto	**12**	+	+		+	
T04024	sotto	**12**	+	+		-	
T15001	dakota	**12**	+	+		+	
T04b001	kenyae	**13**	+	+		+	
T04b012	kenyae	**13**	+	+		+	
T04b054	kenyae	**13**	-	+		+	
T04b060	kenyae	**13**	-	+		+	
T04b073	kenyae	**13**	-	+		+	
T06007	entomocidus	**14**	+	+		+	
T06010	entomocidus	**14**	+	+		+	
T07033	aizawai	**15**	-	+		+	
T07058	aizawai	**15**	-	+		+	
T07180	aizawai	**15**	+	+		-	
T05a030	canadensis	**16**	+	-		-	
T08025	morrisoni	**16**	+	+		+	
CCCT 2259	israelensis	**16**	-	+		+	
T13001	pakistani	**17**	+	+		+	
T13004	pakistani	**17**	+	+		+	

strain	species / serovar	MLST ST	Starch hydrolysis	nhe *B C* genes	nhe *A B C* genes	hbl *C D* genes	hbl *C D A* genes
		Predicted Band size (bp)		*1200*	*1635*	*1900*	*1892*
T13028	pakistani	**18**	+	+		-	
172560 W	*cereus*	**18**		+		+	
WSBC 10249	*cereus*	**19**	+	+	+	+	+
m1280	*cereus*	**19**	+	+	+	+	+
WSBC 10312	cereus	**20**	+	+	+	+	+
WSBC 10277	*mycoides*	**21**	+	+	-	-	-
T09010	tolworthi	**22**	+	+		+	
T09011	tolworthi	**22**	+	+		+	
T09024	tolworthi	**22**	+	+		+	
T09034	tolworthi	**22**	+	+		+	
T09052	tolworthi	**22**		+		+	
T08001	morrisoni	**23**	+	+		+	
T08009	morrisoni	**23**	+	+		+	
T08012	morrisoni	**23**	+	+		+	
T08023	morrisoni	**23**	+	+		+	
T08031	morrisoni	**23**	+	+		+	
NCIB 6349	*cereus*	**24**	+	+		+	+
Cal3	*cereus*	**24**	+	+		+	+
WSBC 10028	*cereus*	**24**	+	+	+	+	+
T05005	galleriae	**25**	-	+		+	
T05033	galleriae	**25**	-	+		+	
T05114	galleriae	**25**	-	+		+	
S 73	*cereus*	**26**	-	+	-	+	-
S 74	*cereus*	**26**	-	+	-	-	-
S 710	*cereus*	**26**	-	+	-	+	-
F3080B/87	*cereus*	**26**	-	+	-	-	-
F3942/87	*cereus*	**26**	-	+	-	-	-
F4810/72	*cereus*	**26**	-	+	-	-	-
F4370/75	*cereus*	**27**	+	1300	+	+	+
F4431/3	*cereus*	**28**	-	+	-	+	+
S 86	*cereus 'terminalis'*	**29**	+	+	+	+	+
S 363	*cereus*	**30**	+	+	-	+	+
S 366	*cereus*	**31**	-	+	+	+	+
ATCC 10987	*cereus*	**32**	-	1300	+	+	+
ATCC 10876	*cereus*	**33**	+	+	+	+	+
ATCC 11778	*cereus*	**34**	+	+	+	-	-
AH621	*cereus*	**35**	+	+		-	-
AH647	*cereus*	**36**	+	+		-	-
AH684	*cereus*	**37**	+	+		-	-
ATCC 4342	*cereus*	**38**	+	1300	-	+	+

strain	species / serovar	MLST ST	Starch hydrolysis	nhe *B C* genes	nhe *A B C* genes	hbl *C D* genes	hbl *C D A* genes
		Predicted Band size (bp)		*1200*	*1635*	*1900*	*1892*
SPS 2	*cereus*	**39**	-	+	+	+	+
TSP 11	*cereus*	**40**	+	+		+	+
WSBC 10202	*weihenstephanensis*	**41**	+	-	-	-	-
WSBC 10364	*weihenstephanensis*	**42**	+	-	-	-	-
m1278	*cereus*	**43**	+	+	+	+	+
m1292	*cereus*	**44**	-	+	+	+	-
m1293	*cereus*	**45**	-	-	-	-	-
m1550	*cereus*	**46**	-	+	+	+	+
m1576	*cereus*	**47**	-	-	+	+	+
T01246	thuringiensis	**48**	+	+		+	
T04236	sotto	**49**	+	+		+	
T05a001	canadensis	**50**	+	+		+	
T05a015	canadensis	**51**	+	+		+	
T05a019	canadensis	**52**	+	+		+	
T07146	aizawai	**53**	-	+		+	
T07196	aizawai	**54**	+	+		+	
T10016	darmstadiensis	**55**	+	+		+	
T10001	darmstadiensis	**56**	+	+		+	
T10003	darmstadiensis	**56**	+	+		+	
T10017	darmstadiensis	**56**	+	+		+	
T10018	darmstadiensis	**56**	+	+		+	
T10024	darmstadiensis	**57**	+	+		-	
T15006	dakota	**58**	-	+		+	
T18001	kumamotoensis	**59**	-	+		+	
T18002	kumamotoensis	**59**	-	+		+	
T18004	kumamotoensis	**60**	-	-		-	
T05125	galleriae	**61**	-	+		-	
T08019	morrisoni	**63**		+		+	
T08003	morrisoni	**64**		+		-	
T14531	israelensis	**65**		+		+	
m1552	*cereus*	**66**	-	-	+	+	+
T08121	morrisoni	**67**		+		+	
T09016	tolworthi	**68**		+		+	
T09043	tolworthi	**68**		+		+	
T08016	morrisoni	**69**		-		+	
T08051	morrisoni	**70**		+		+	
T09020	tolworthi	**71**		+		+	
R 3238/03	*cereus*	**72**		+		+	
R 3149/03	*cereus*	**73**		+		+	
R 2955/03	*cereus*	**73**		+		+	

strain	species / serovar	MLST ST	Starch hydrolysis	nhe *B C* genes	nhe *A B C* genes	hbl *C D* genes	hbl *C D A* genes
		Predicted Band size (bp)		*1200*	*1635*	*1900*	*1892*
R 3098/03	*cereus*	**74**		+		-	
R 3039/03	*cereus*	**75**		+		+	
191560 K	*cereus*	**76**		-		-	
168287 M	*cereus*	**77**		-		-	
S79	*cereus 'fluorescens'*	**79**	+	+	+	-	+
TSP 8	*cereus*	**80**	-	-	-	-	-
S78	*cereus 'albolactis'*	**81**	-	+		+	+
K8	*cereus*	**82**	+	+		+	-
WS 3118	*pseudomycoides*	**83**	-	-	+	-	+
TSP 10	*cereus*	**86**	+	+		+	+
WSBC 10360	*mycoides*	**87**	-	-	-	-	-
STR4	*mycoides*	**88**		+		+	
Number of negative results			38	14	17	32	18
Total number of strains tested			108	128	39	128	49

Strains which are negative for both *nheBC* and *hblCD* enterotoxin genes are shaded.

Bacillus thuringiensis strains are designated by serotype (not italicized)

*Ames. The results for the enterotoxin genes from *B. anthracis* were obtained by a BLAST search of the *B. anthracis* sequence at TIGR.

Some of the enterotoxin PCR results from *Bt* strains were carried out by A. Gaviria Rivera, as part of her Ph. D thesis and are included here for comparative purposes.

Table 3.30 Strains analysed for starch hydrolysis and enterotoxin genes, but not included in MLST.

strain	species / serovar	MLST ST	Starch hydrolysis	nhe *B C* genes	nhe *A B C* genes	hbl *C D* genes	hbl *C D A* genes
		Predicted Band size (bp)		*1200*	*1635*	*1900*	*1892*
S61	*cereus*		-	+	-	-	-
S62	*cereus*		+	+	+	-	-
S63	*cereus*		+	+		+	+
S67	*cereus*		+	+		+	+
S68	*cereus*		+		+	+	+
S72	*cereus*		+	1300	-	+	+
S76	*cereus*		+	+		+	+
S77	*albolactis*		-	+		-	-
AH517	*cereus*		+	+		+	
AH563	*cereus*		+	+		+	
AH581	*cereus*		+	+		+	
AH636	*cereus*		+	+		-	
AH651	*cereus*		+	+	+	-	+
AH656	*cereus*		+	+		+	-
AH662	*cereus*		+	+		-	-
AH671	*cereus*		+	-	-	-	+
AH681	*cereus*		-	+		-	-
AH685	*cereus*		+	+		+	-
AH686	*cereus*		+	+		-	-
AH691	*cereus*		+	+		-	-
AH693	*cereus*		+	+		-	-
WS2629	*thuringiensis*		+	+	+	+	+
WS2734	*thuringiensis*		+	+		+	+
WSBC28002	*thuringiensis*		+	+	+	+	+
H5	*cereus*		-	+	+	-	-
BMVIII	*cereus*		+	+		+	+

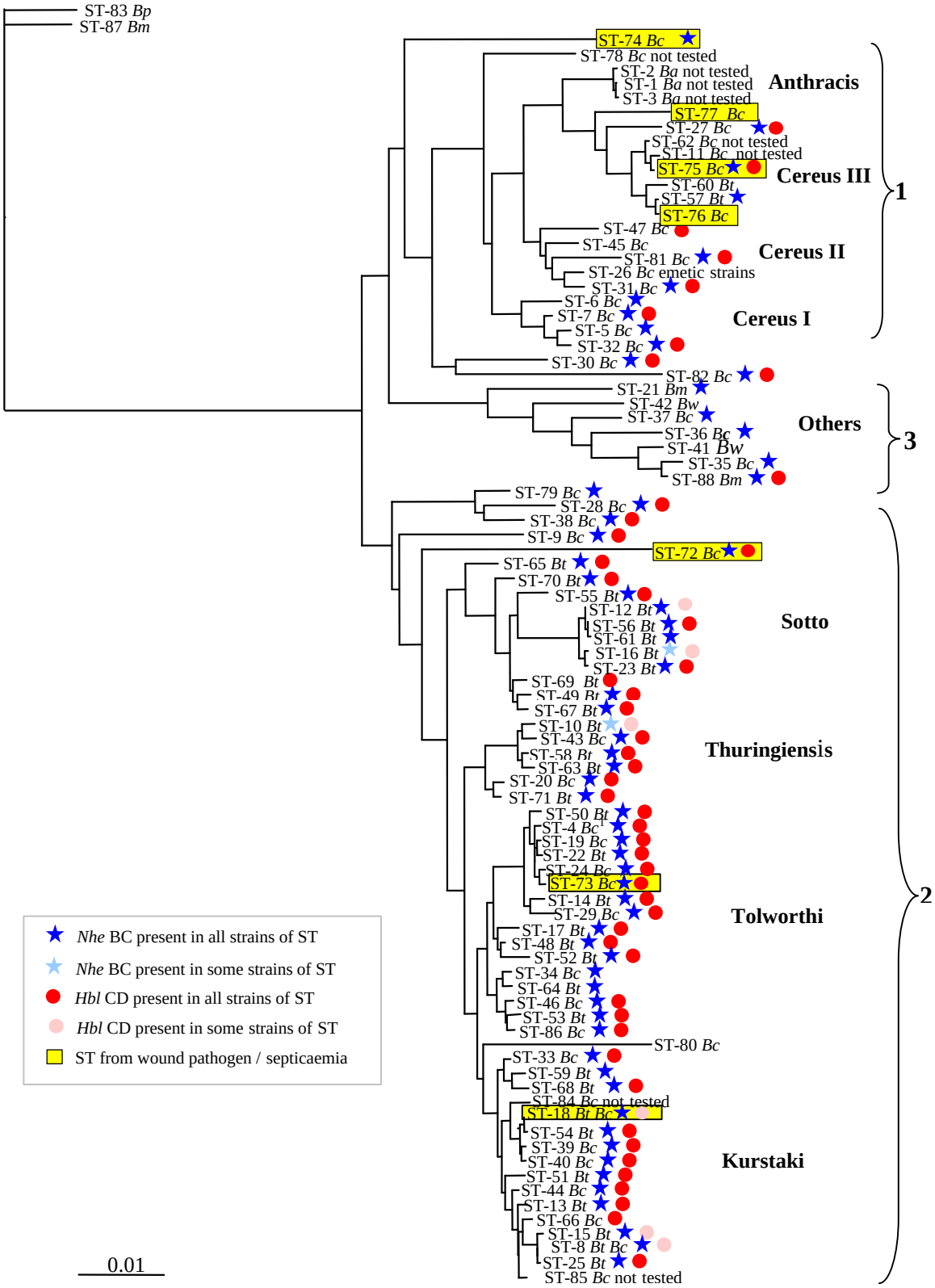


Figure 3.62 Rooted NJ tree of 88 STs (147 strains), showing the occurrence of the non-haemolytic (*nhe* - blue star) and the haemolytic (*hbl* – red circle) enterotoxin genes. STs of strains isolated from wounds and septicaemia are highlighted in yellow (Barker *et al.*, 2005).

Chapter 4 DISCUSSION

An MLST database for the *B. cereus* group has been developed and is now accessible at http://pubmlst.org. Analysis of the seven housekeeping genes selected for MLST has given new insight into the evolutionary history and phylogenetic structure of this diverse group of bacteria.

> *'I think you have to base evolutionary relationships on the DNA sequence. It's the most fundamental way of classifying organisms. Any other type of data essentially represents some sort of phenotype that relates back to the DNA sequence'.*
> (T. Read, interview 2003).

The changes accumulated in the seven genes resulted in some amino acid changes but most of these were synonymous. Very rarely, an amino acid change altered the protein structure; one instance of this occurring in the *ilvD* gene is shown in Figure 3.5. Consequently, by using concatenated nucleotide sequence data for most of the phylogenetic inference, all base changes in the DNA sequence were considered and this is generally believed to be the best method available for comparing closely related bacteria where recombination is limited (Priest *et al.*, 2004; Aanensen & Spratt, 2005).

Species ?

Representatives of the six 'species' comprising the *Bacillus cereus* group have been examined in this work and it was apparent that separation into *B. anthracis*, *B. cereus* and *B. thuringiensis* (with *B. mycoides*, *B. pseudomycoides* and *B. weihenstephanansis* as additional 'species') was inconsistent with my results. The original criteria used to define the six separate species were based on phenotypic properties and pathogenic traits. However, the difficulties in defining bacterial species using phenotypic criteria have long been acknowledged: -'*the component strains of many species are relatively heterogeneous and form a graduated series of differences like the gradations of a spectrum in which colors fade into one another in sequence*' (Gordon *et al.*, 1973).

The concept of the species was developed to separate organisms into similar groups of similar status. The 'Biological Species Concept' was developed by Mayr in 1944 and defined species as '*….groups of actually or potentially interbreeding natural*

populations which are reproductively isolated from other such groups'. Later, in 1970, Mayr revised this definition, to include an ecological component; *'…..a reproductive community of populations (reproductively isolated from others) that occupies a specific niche in nature.'* (Skelton, 1993).

Since this time, how species are defined has been the subject of lively debate and there are three main ideas: - evolutionary species, biological species and recognition species

'Evolutionary species'. A species comprises individuals with similar evolutionary history. This idea is fine, but all individuals are different and there is some difficulty in deciding on the degree of similarity required to define a single species. The borderlines between such species are decidedly hazy.

'Biological species'. A species is made up of individuals that can successfully reproduce with each other. In this concept, species are isolated from other species by barriers to reproduction such as requiring a different habitat, season, physiological compatibility or requiring particular courtship displays. There are practical difficulties with these ideas in that they are not applicable to asexual organisms, including prokaryotes. Even in sexual animals there are unclear areas concerning the degree of reproductive isolation required to consider organisms as separate species (Skelton, 1993). Nevertheless, this is the most commonly accepted definition of a species.

'Recognition species'. A species consists of two parent organisms, which share a common fertilization system. Again, this definition is inapplicable to asexual organisms.

It is apparent that among eukaryotic organisms there are difficulties defining species. When the emphasis turned to reproductive capability, the definition of a prokaryotic species became more of a problem, as these organisms are basically asexual.

The species concept applied to bacteria

As phenotypic methods developed, computer based methods of analyzing the data led to numerical taxonomy, which allowed large numbers of bacteria to be compared for many phenotypic characters (Senath & Sokal, 1973). Similar individuals were found to cluster together in groups and phenetically defined species could be delineated in dendrograms. For example, in an extensive numerical taxonomy study of *Bacillus, s*pecies such as *B. subtilis* and *B. licheniformis* were well defined phenotypically, but strains of *B. cereus* and *B. thuringiensis* were less obvious (Priest *et al.,* 1988).

A breakthrough in bacterial systematics occurred with the development of whole genome DNA-DNA hybridization and the discovery that strains which clustered together by phenotypic analysis tended to share >70% DNA hybridization. A further refinement in identifying bacterial species uses DNA sequence divergence of 16S rRNA genes and a difference of >3% implies that the strains belong to separate species (Cohan, 2002).

Using phenotypic criteria and the basis of shared characters, *B. anthracis, B. cereus, B. thuringiensis* and the rhizoid forms are best defined as one species despite their different pathogenic properties (Gordon *et al.,* 1973). Moreover, studies of DNA hybridization confirmed these phenotypic observations (Kaneko *et al.,* 1978) and a MEE study of *B. anthracis, B. cereus* and *B. thuringiensis* strains suggested that the group be considered one species, with *B. anthracis* as a lineage of *B. cereus* (Helgason *et al.,* 2000 b).

However, arguments against a rigid definition of 'species', depending on factors such as 70% DNA hybridization or 97% 16S rRNA homology have been raised (Roselló-Mora, 2003). Roselló-Mora suggested that the term 'species' does not need to be rigid, but may be dependent on the organisms being examined, even when plants or animals are being considered. Such a concept is not new:

'A variety of species concepts are necessary to adequately capture the complexity of variation patterns in nature……we must resist the urge to superimpose false simplicity. If 'species situations' are diverse, then a variety of concepts may be necessary and desirable to reflect this complexity' (Mishler & Donoghue, 1982).

The recent concept of 'fuzzy' species for bacteria which are highly recombinogenic emphasizes that even when species clusters can be resolved from MLST data, intermediate strains occur and the species boundaries are probably best described as fuzzy (Hanage *et al.,* 2005).

One way forward may be to adopt a flexible polyphasic approach to bacterial taxonomy, as suggested by Gillis *et al.,* (2001). Such an approach could be based on DNA-DNA hybridization, 16S rRNA gene sequence similarity and additional standardized genomic, chemotaxonomic and phenotypic tests as appropriate. The relatively new technique of MLST could play a major role in such a scheme. Additionally, examination of the micro-evolution of clonal groups/lineages could be carried out by sequencing more polymorphic genes, such as those coding for virulence, antibiotic resistance, or cell surface proteins which tend to evolve more rapidly than genes which control metabolism. At the other extreme, MLST schemes could be developed to cover whole genera if a set of common 'core' genes could be identified (Cooper & Feil, 2004). All these data could be linked together and made readily available electronically.

Population structure in bacterial species.

If we accept the polyphasic definition of bacterial species, we can consider the short-term evolutionary patterns of the population of strains that compose that taxon. Bacterial populations evolve over a spectrum of models defined by two extremes:

(1) Diversity reduction and periodic selection (clonal).
(2) Recombinationaly active (panmictic).

(1) Diversity reduction and periodic selection

Speciation in asexual organisms such as bacteria has been addressed in the 'Cohesion Species Concept' (Templeton, 1989). Genetic diversity within a species is limited by cohesive forces, such as natural selection, which happen when a bacterium adapts by mutation to a new habitat or becomes able to utilize a new resource, as illustrated in Figure 4.1.

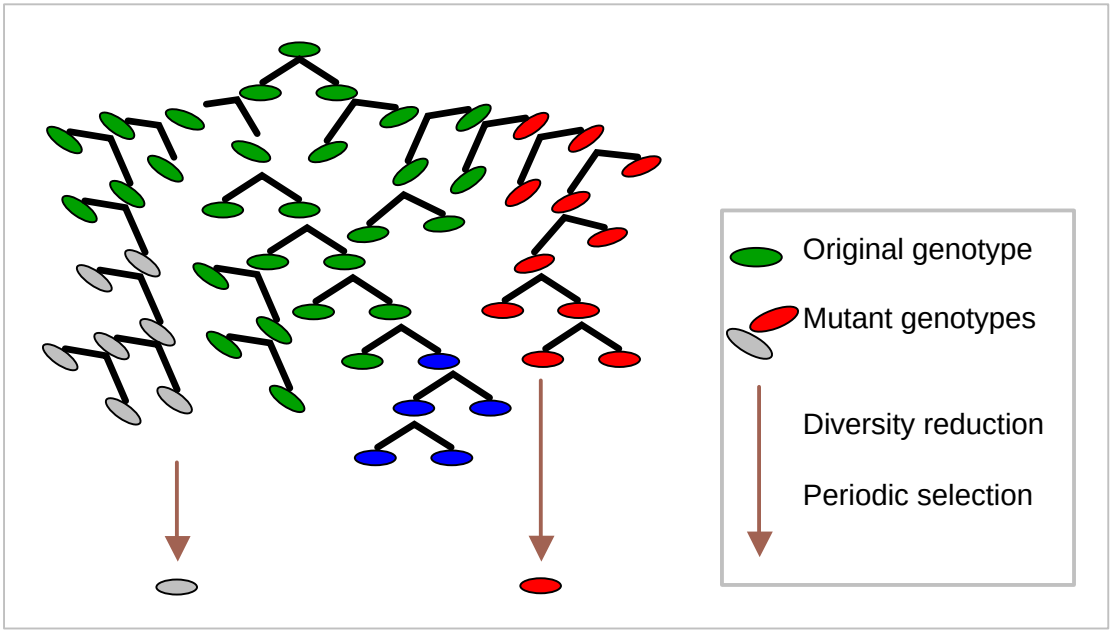


Figure 4.1 Clonal population structure – periodic selection.
The original bacterium (top green oval) divides and mutant strains arise naturally.
Successful mutants survive and other strains become extinct (Adapted from Maiden, 1993)

The progeny from the successful bacteria rapidly increase at the expense of the unadapted strains until the mutant strain becomes 'fixed' (100% present). Later, the environment may alter and another mutation will produce successful progeny, eliminating the rest of the population. Such a cyclic process is called periodic selection. Bacteria attain separate species status when an adapted population is unaffected by another mutation or adaptation in a separate population. Both types of population are fit ecotypes and can survive together (Cohan, 2002).

In a phylogeny derived from DNA sequence data, different ecotypes will emerge as cluster groups or clonal complexes, with an identifiable common ancestor (this will be clearer in species with little or no recombination) (Cohan, 2002).

(2) Recombinationaly active bacteria

Horizontal gene exchange in bacteria is rare and is not necessarily limited to members of the same species. Gene transfer may confer a new metabolic function on a bacterium, creating a novel species almost instantly (Figure 4.2).

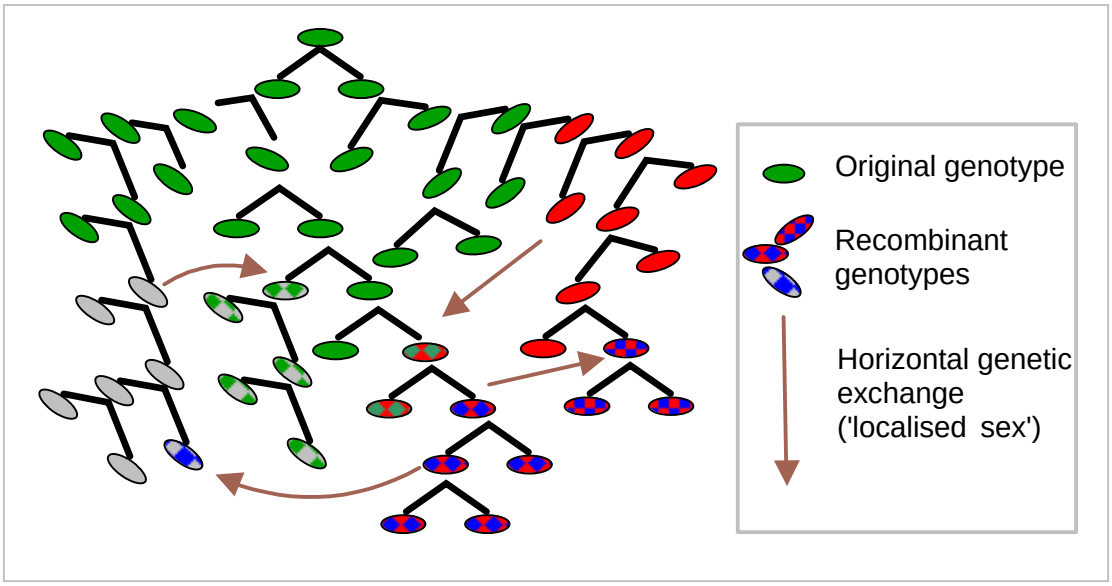


Figure 4.2 Population with horizontal gene transfer (recombination)
Exchange of genetic material across the population allows new phenotypes to arise without the need for any naturally occurring mutations (Adapted from Maiden, 1993)

Bacterial species are usefully defined as groups of strains in which over 70 % of their DNA reassociates and the thermal stability of the reassociated DNA is within 4°C of that of the homologous DNA (Dykhuizen & Green, 1991). A related criterion for species definition is the extent of horizontal gene transfer within the group of strains being examined. Groups of strains that recombine with each other, but not with strains from other groups, constitute a species. Unfortunately for this idea, bacteria are able to transfer genetic material between distant species, albeit rarely. However, recombination is much more common within species than between species. (Cohan, 2002). Indeed, these two concepts are inter-related. Horizontal gene exchange commonly relies on homologous recombination of the incoming DNA. This is much more likely to occur in closely related strains (i.e. strains which display > 70 % DNA hybridization)

Clonal and panmictic populations can be distinguished using MLST data. Clonal groups, which are diversifying due to a gradual increase in point mutations, will initially

have very similar DNA sequences and therefor the MLST allelic profiles will also be similar (Feil *et al.,* 2003). Additionally, the loci that have mutated will have very few nucleotide differences from the original sequence, perhaps only one nucleotide difference. If recombination has occurred, alleles are likely to have many nucleotide differences as a section of DNA sequence will have been acquired from elsewhere - from a sister strain, generally by transduction in non-competent species.

Estimates of recombination within lineages

The extent of recombination may be estimated by an examination of single locus variants (SLVs) within lineages or clonal complexes, as demonstrated for *S. aureus* by Feil *et al.,* (2003). The *S. aureus* data were analyzed using BURST (Based Upon Related Sequence Types) analysis, which splits data into clonal complexes (Spratt *et al.,* 2004). A clonal complex is generally defined as a group of strains that share five or more alleles with one other member of the group. The BURST program can be used to derive clonal complexes from an MLST dataset and to assign a putative ancestral type. For *S. aureus*, eight clonal complexes were found and 35 SLVs. Only one SLV was deduced to have probably arisen by recombination rather than mutation. To estimate recombination, the criteria applied to the SLVs were that the allele should differ from the ancestral allele (preferably at more than one nucleotide site) and that the variant allele should be present in other strains in the dataset.

The *B. cereus* data did not resolve into clonal complexes, probably due to high diversity. However, analysis using BURST grouped those STs that differed by only a single locus (Table 4.1). From nine lineages, ten single locus variants (SLVs) were identified, four occurred in the Sotto lineage, two were found in Kurstaki and a single SLV was present in Tolworthi, Cereus I and Cereus III. There were no SLVs in Thuringiensis, Cereus II or the 'Others'. As in the *S. aureus* data, only one SLV could be confidently assigned as having arisen from a recombination event. In the Sotto lineage, ST-56 differed from ST-55 in the *ilvD* allele; *ilvD-7* and *ilvD*-8 both with two amino acid changes (the amino acid change from threonine to methionine resulted in a strand changing into a helix (Figure 3.5)). The variant allele, *ilvD-8*, was present in four other STs. Interestingly, three of the STs containing *ilvD-8* were not members of the Sotto lineage but were assigned to Tolworthi and the fourth ST (ST-9) was unassigned, strongly supporting the involvement of horizontal gene transfer (HGT).

Table 4.1 Detection of recombination by examination of Single Locus Variants (SLVs).

Variant alleles within SLVs from Lineages

Lineage	ST	ST of SLV	Variant locus in SLV	Allele change	Other STs with SLV allele	No of nucleotide differences	Nucleotide changes	Amino acid changes	Mode
Anthracis	1	3	*glp*	1 to 2	0	1	AAC to ACC	N to T	mutation
Cereus I	7	32	*tpi*	7 to 16	0	7	ACA to ACT GAT to GAC GAA to GAG GAA to GCT ATC to ATT TAT to TAC GGG to GAG	none none none E to A none none G to E	mutation or recombination
Cereus II	No single locus variants								
Cereus III	62	75	*glp*	38 to 44	0	5	TTG to TTA TTA to CTA GCA to GTA GGG to GGC ATC to ATT	none none A to V none none	mutation or recombination

Table 4.1 (continued from previous page)

Lineage	ST	ST of SLV	Variant locus in SLV	Allele change	Other STs with SLV allele	No of nucleotide differences	Nucleotide changes	Amino acid changes	Mode
Sotto 1	56	12	*pyc*	26 to 10	67	1	CAA to CAG	none	mutation
	56	55	*ilv*	7 to 8	9, 29, 4, 22	29	numerous	D to E T to M	recombination
	56	61	*pta*	2 to 27	0	1	ACT to AGT	T to S	mutation
Sotto 2	23	16	*pur*	5 to 6	65	2	TGT to CGT CCA to TCA	C to R P to S	mutation or recombination
Thuringiensis	No single locus variants								
Tolworthi	34	86	*pur*	2 to 8	29, 46	7	ACT to ACG CGC to CGT AAA to AAG ACG to ACT AAG to AAA GCG to GCA AAG to AAA	none	mutation or recombination
Kurstaki 1	8	25	*glp*	8 to 30	0	1	GGT to GAT	G to D	mutation
Kurstaki 2	18	54	*pta*	12 to 32	0	1	CTT to CAT	L to H	mutation
Others	No single locus variants								

In general, alleles in the *B. cereus* group appear to be about ten times more likely to change by point mutation than by recombination. For *S. aureus*, alleles were estimated to be 15 times more likely to change by point mutation than by recombination. The figures for *B. cereus* and *S. aureus* contrast with the estimated rates for the naturally transformable species *N. meningitides* and *S. pneumoniae*, where allele change by recombination is 5 to 10 times more likely than mutation (Feil *et al.*, 2000).

Recombination and MLST gene trees

An examination of individual gene trees also provides a mechanism to determine the extent of gene transfer. If there has been no horizontal gene transfer trees will be congruent, but if genes have been transferred the topology of the trees will vary (Feil & Spratt, 2001). In a MLST study of *S. aureus*, an analysis of gene trees from 25 representative STs was used to investigate congruence between loci. One particular allele, *arcC*, resulted in a different tree from the others, and an inspection of the genome sequence showed that this allele was close to putative virulence factors which are subject to diversifying selection. The *arcC* allele may have 'hitch-hiked' recombinational events associated with the virulence genes. However, the overall structure of *S. aureus* was found to be mainly clonal and due to point mutations with limited effects related to recombinational events as indicated above (Feil *et al.*, 2003).

In this project, individual gene trees were constructed for a representative set of 21 STs from the *B. cereus* group (Section 3.1.6.i). Examination of the seven trees revealed that while most of the trees were broadly similar in structure, there was evidence for some gene exchange at all loci. The alleles found to be most recombinogenic were *glpF* and *ilvD* as can be seen when Figures 3.20 and 3.22 are compared with the concatenated tree in Figure 3.19. Split decomposition analysis of *glpF* and *ilvD* (Figures 3.28 and 3.30) supported the observation that ST-83, *B. pseudomycoides*, was not an 'outlier' for those alleles whereas in the concatenated tree and the other five allele trees, ST-83 was a definite outlier. Additionally, for the *glpF* gene, a new cluster group was formed, comprising ST-8 (*Bt* kurstaki), ST-32 (ATCC 10987) from Cereus I and ST-38 (ATCC 4342) which was unassigned. Interestingly, for *ilvD* the three strains from clade 3 were dispersed on the NJ tree, suggesting recombination in *ilvD*. While the five other trees

agreed broadly with the concatenated tree, there was an interesting grouping in the *pur* allele (Figure 3.24), ST-18 from the Kurstaki lineage of clade 2 (*Bt* pakistani and a *B. cereus* wound pathogen, 172560 W), joined two STs from clade 1, ST-26 (five emetic toxin strains) and ST-32 (ATCC 10987) in a cluster.

ST-21, *B. mycoides*, was associated with clade 3 for most of the trees, but in the *gmk* tree ST-21 was in clade 2 and in the *ilvD* tree, ST-21 was recovered in clade 1. Other STs also transferred to different clades depending on the allele examined. For example, ST-38 (ATCC 4342), moved between clade 1 and clade 2 and ST-82 (strain K8 from a cheese dairy) transferred from clade 1 to clade 3 in the *glpF* and *gmk* trees. However, the overall structures of the individual gene trees were sufficiently similar to suggest a weakly clonal structure for the *B. cereus* group. This was supported by an analysis of congruence of ML trees from concatenated sequence data and from individual loci, using strains that were included in this project (Priest *et al.,* 2004). This conclusion is consistent with the BURST analysis, which suggested a 10-fold excess of mutations over recombination.

Horizontal gene exchange is relatively rare within the *B. cereus* group and the mode of action is likely to be plasmid or phage driven since unlike *B. subtilis*, strains of *B. cereus* are not naturally competent for DNA mediated transformation (Dubnau, 1991; Cohan, 2002).

The index of association (I_A) is a statistical test used to quantify the extent of clonality in bacterial (Maynard Smith *et al.,* 1993). The indexes of association for several bacterial species are shown in Table 4.2. The I_A for the *B. cereus* group dataset was 2.84 for the whole dataset and 1.74 when single STs were used (section 3.1.7.ii), indicating a weakly clonal population. Another MLST study of the *B. cereus* group (Helgason *et al.,* 2004), found an I_A of 3.54 when using all 77 isolates and 1.90 when only one representative of each ST was used. These figures for I_A from two independent studies agree well and provide strong confirmation of a weakly clonal population with limited recombination. This can be compared to the population structure for *Neisseria meningitidis*, a species that is known to be naturally transformable and subject to recombination (Maiden, 1993). A study of 107 strains of *N. meningitidis* found strong evidence for recombination (Holmes *et al.,* 1999).

Table 4.2 Index of Association (I_A) for representative bacterial species

Species	Number of strains	I_A all isolates	I_A STs only	Population	Reference
B. cereus	146	2.84	1.74	Weakly clonal	This thesis, 2005
B. cereus	77	3.54	1.90	Weakly clonal	Helgason *et al.*, 2004
C. difficile	29	5.17	2.01	Clonal	Lemée, *et al.*, 2005
C. jejuni	194	2.02	0.57	Weakly clonal	Dingle *et al.*, 2001
N. meningitidis	107	4.57	0.65	Epidemic	Holmes *et al.*, 1999
P. aeruginosa	143	0.29	0.17	Non clonal	Curran *et al.*, 2004

When the whole dataset of 107 strains of *N. meningitidis* was tested, including invasive strains with the same ST, an I_A of 4.57 was found, but when single STs were examined, the I_A was reduced to 0.65. Such a big change in I_A is characteristic of an epidemic type of population (Holmes *et al.*, 1999). Even if a bacterial population is non clonal, a sub population may acquire genes from the environment or by horizontal gene transfer and become successful, giving rise to an epidemic clonal structure, which may be short-lived (Figure 4.2). Further work on 126 isolates of *N. meningitidis* recovered 10 clonal complexes and found evidence for extensive HGT as judged by numerous nucleotide changes in the same allele and the appearance of variant alleles which are also present elsewhere in the dataset). The likelihood of a single nucleotide site changing by recombination rather than mutation (the r/m rate) estimated that recombination was 80 times more likely than mutation in *N. meningitidis*, which is twice the r/m rate estimated for *E. coli* (Feil *et al.*, 1999).

Many MLST studies analyse the data using allelic profiles. This approach means that the magnitude of the DNA sequence changes may be lost, because a new allele number could represent a single base change or many. An earlier MLST study of the *B. cereus* group (Helgason *et al.*, 2004) divided each allele into three segments to increase the resolving power, but by using concatenated sequence data, all base changes are utilized. A recent MLST study of 770 strains of 11 *Neisseria* species used concatenated sequence data rather than alleles to compare the data. The *N. meningitidis* tree (from the concatenated sequence data) resolved the three major species, *N. lactamica*, *N.*

gonorrhoeae and *N. meningitidis*, but the boundaries were fuzzy, due to intermediate types (Hanage *et al.*, 2005). The use of single gene trees for the *N. meningitidis* data failed to resolve the separate species due to the effects of recombination.

In this thesis, trees constructed from the concatenated sequence data for *B. cereus* recovered three main clusters that we have labelled clades. Clade 1 was predominantly *B. cereus* strains and included *B. anthracis*. Clade 2 was a heterogeneous mixture of *B. cereus* and *B. thuringiensis* strains. One possibility to define the species would be to restrict the name *B. cereus* to isolates that fall into clade 1. Unfortunately this would include some *B. thuringiensis* species and *B. anthracis*. Such an approach has been suggested for the highly recombinogenic pneumonococcus. Trees constructed using concatenated sequence data showed that the data separated into two groups: - one of the groups clustered with the serotypable *Streptococcus pneumoniae* while the other group contained more diverse strains (Hanage *et al.*, 2005). From this study, the authors proposed that *S. pneumoniae* should be re-defined to include only the strains falling into cluster 1. However, this approach would not be helpful when identifying pathogenic potential as some of the cluster 2 isolates were from children with ear infections. Likewise, in *B. cereus*, some isolates from wound infections were recovered in clade 2 and a few crystalliferous *B. thuringiensi*s strains were recovered in clade 1 (Priest *et al.*, 2004, Barker *et al.*, 2005).

Clonal groups

Although evidence of some recombination was found in the *B. cereus* group, several clusters or clonal groups were also recovered, e.g. *B. anthracis*, the emetic toxin producing strains in clade 1 and lineage Sotto in clade 2. The other MLST study of the *B. cereus* group also recovered some clonal groups e.g. all five *B. anthracis* strains assigned to ST-1 and nine *B. cereus* strains isolated from periodontitis comprised ST-38 (Helgason *et al.*, 2004). Some other species of bacteria that have been studied using MLST, such as *Mycobacterium* species and *Yersinia* species also formed clonal groups, showing very little variation in the sequences of their housekeeping genes (Lan & Reeves, 2000; Lan & Reeves, 2001; Cohan, 2002). Clustering derived from DNA sequences results in clusters, sub-clusters and sub-sub clusters and the status of species, clade, clone or ecotype must be determined. In some cases, the limited genetic diversity within a species makes assignment of isolates to clonal complexes relatively simple

(Feil *et al.,* 2003). While this approach was successful with *Campylobacter* species, *N. meningitidis* and *S. aureus,* the *B. cereus* group is genomically diverse and most isolates did not resolve into clonal complexes using BURST.

BURST analyses the data on the basis of allelic profiles rather than nucleotide sequences. The genomic diversity present in the *B. cereus* group is reflected in the relatively high average number of alleles / locus: - 40 alleles/locus from 146 strains (88 STs). For *C. jejuni* the average number of alleles/locus was 36 alleles from 194 isolates (155 STs) (Dingle *et al.,* 2001) and for *N. meningitidis* an average of 21 alleles/locus from 107 isolates (49 STs) (Maiden *et al.,* 1998). The initial *S. aureus* study recovered an average of 14 alleles/locus from 155 isolates (53 STs) (Enright *et al.,* 2000) and a more recent study of *S. aureus* recovered only 75 STs from 334 isolates. Although the number of alleles/locus was not reported for the last mentioned study, the data were analysed using the BURST algorithm for clonal complexes (Feil *et al.,* 2003). There were 75 STs in the dataset and these clustered, using BURST, into eight major clonal complexes and three minor complexes. The influence of recombination on the phylogeny of *S. aureus* was estimated to be negligible, in contrast to *N. meningitides* as mentioned earlier. For *S. aureus*, a nucleotide site was estimated to be 15 times more likely to change by point mutation than by recombination and the population structure of *S. aureus* was found to be highly clonal, with strains from disease being found in all clonal complexes. In the context of *S. aureus* (clonal) and *N. meningitidis* (non-clonal), the *B. cereus* group may be defined as intermediately clonal.

However, *S. aureus* and the *B. cereus* group do have an important feature in common. Wound pathogens were not restricted to a particular group but were isolated from several clonal complexes or lineages, as shown for *B. cereus* in Figure 3.62.

Due to the diversity of the *B. cereus* data, clonal complexes were not resolved by the BURST algorithm, with the interesting exception of the clonal Sotto lineage in clade 2. The Sotto lineage comprises several closely related entomopathogenic *B. thuringiensis* serovars such as *Bt* darmstadiensis *Bt* israelensis, *Bt* morrisoni, and *Bt* sotto. ST-56 (four strains of *Bt* darmstadiensis) was assigned as the founder sequence type. The strains within Sotto had three alleles in common: - *glp-15*, *gmk-7* and *tpi-13*. Additionally, *pta-2* was found in all but one of the Sotto strains. Sotto is an example of

a highly successful niche-adapted bacterium with a virulent entomopathogenic lifestyle and this may be the reason why a clonal complex has evolved. It may be expected that when the number of strains in the *B. cereus* database has increased more clonal complexes will be revealed. An extensive MLST study of 912 methicillin resistant and methicillin susceptible strains of *S. aureus* found that methicillin resistant clones of *S. aureus* (MRSA) have arisen several times from successful clones of methicillin susceptible *S. aureus* (Enright *et al.,* 2002), either by horizontal transfer of the *mec* gene which confers resistance, or by descent as a single locus variant, from an existing MRSA clone. For the *B. cereus* population as with *S. aureus*, isolates with the same ST probably have different virulence genes.

Traditional criteria for classifying the *B. cereus* group at the sub-specific level have included serotyping, host range, niche adaptation and plasmid content. It is therefore of interest to compare MLST groups with these traditional typing schemes.

Serotyping

In analysis of the *B. thuringiensis* data, strains with the same 'H' serotype were found to have different STs and were often recovered in different lineages as shown in Figure 3.9 (serotypes for the *B. thuringiensis* strains are listed in Appendix A). In particular, strains from *Bt* morrisoni (serotype 8a, 8b) were found in lineages Sotto, Thuringiensis and Tolworthi. *Bt* canadensis (serotype 5a, 5c) strains were recovered in all four lineages of clade 2 and while *Bt* darmstadiensis strains were normally recovered in lineage Sotto (ST-55 and ST-56), a single atypical strain of *Bt* darmstadiensis (ST-57, T10024 isolated in Pakistan) was recovered in lineage Cereus III of clade1. Similarly, two strains of *Bt* kumamotoensis were assigned to lineage Kurstaki while another *Bt* kumamotoensis strain (T18004, isolated in Iraq) was recovered in Cereus III and clustered close to the atypical *Bt* darmstadiensis mentioned above. Only *Bt* israelensis (serotype 14) and *Bt* sotto (serotype 4a, 4b) were confined to lineage Sotto and not represented in other lineages. *Bt* israelensis is generally regarded as a clonal group characterised by the possession of the dipteran specific *cry4* genes (Priest *et al*., 1994).

Previous work has found some correlation between serovar and genomic typing. Characterization of 10 serovars by ribotyping found a strong correlation between ribotype and serotype for the serovars aizawai, galleriae, kurstaki, morrisoni,

thuringiensis and tolworthi. In addition, *Bt* tenebrionis was distinguishable from *Bt* morrisoni by ribotyping (Priest *et al.*, 1994). Another investigation, involving the characterization of 126 strains of *B. thuringiensis* from 57 serovars by RAPD typing, found that several of the serovars were genomically homogeneous, e.g serovars thuringiensis and kurstaki, but other serovars were heterogeneous e.g. serovars canadensis and aizawai. (Gaviria *et al.*, 2002). A genetically heterogeneous background was suggested with plasmid DNA exchange and associated insect toxicity giving rise to successful clonal groups. This theory was supported by a MEE comparison of 154 *B. cereus / thuringiensis* soil isolates which were not necessarily insect pathogens (Helgason *et al.*, 1998). No correlation between serotype and MEE genotype could be found among the largely non-pathogenic strains and no clonal population structure was recovered. These strains recovered from soil may be the panmictic background population, from which clonal groups emerge, adapted to fill an available niche associated with Cry protein production.

In this more detailed and comprehensive analysis of the correlation between serovar and genomic type the relationships have been less clear. Some serovars were genomically homogeneous, for example, seven strains of *Bt* israelensis were compared by RAPD (Figure 3.42) and only one gave a slightly different pattern, with an extra band. MLST analysis subsequently found two different alleles in this strain (T14531, isolated in Vietnam), but pathogenicity to mosquito larvae was retained (Table 3.28). Two strains from other serovars, T05a030 (*Bt* canadensis) and T08025 (*Bt* morrisoni) had the same RAPD pattern as CCCT 2259 (*Bt* israelensis) (Figure 3.42). These strains had the identical sequence type (ST-16) and all three amplified the dipteran-specific *cry4* gene, thus isolates from three different serotypes formed a clonal group with similar pathogenicity to insects. Five strains of *Bt* morrisoni with RAPD pattern 3/3a typical of *Bt* israelensis were found to have the *cry4* gene, again suggesting a close relationship with *Bt* israelensis.

The presence of *cry1* and *cry1b* genes was found in 18 out of 20 strains of *Bt* thuringiensis and agreed with the ST-10 designation. Additionally, a variant strain of *Bt* thuringiensis, (T01246 from Iraq) did not amplify the *cry1/1b* genes and was assigned to ST-48, with no alleles in common with the other strains of *Bt* thuringiensis. *Bt*

morrisoni strains with *cry1* usually gave RAPD pattern 1 and ST-23, with one exception, strain T08016 which had a different *ilvD* allele.

A lack of correlation between serotyping schemes and MLST has been encountered in studies of other bacteria, as exemplified by an investigation into *Campylobacter jejuni* (Dingle *et al.,* 2001). Some clonal lineages found in *C. jejuni* were associated with the Penner serotyping scheme, based on heat-stable antigens, but often strains with a particular serotype were found in more than one clonal lineage. *C. jejuni,* isolates from human disease clustered together while isolates from poultry were broadly distributed. This study also compared sequences for two cell surface components and found that the rapidly evolving cell-surface genes were poor indicators of clonal groupings. Serotyping is also an unreliable indicator of insect pathogenicity (*cry* gene content) particularly for *Bt* serovar morrisoni (Table 3.25). Moreover, many *B. cereus* strains react with the flagellar 'H' antisera devised for *B. thuringiensis* (Shisa *et al.,* 2002). A similar problem arose in a study of enteroinvasive *E. coli* (EIEC) strains, which are usually identified by 'O' antigen serotyping, but several of the 'O' antigens used are also present in *Shigella* species (Lan *et al.,* 2004).

In this study, MLST and RAPD typing proved to be better indicators of *cry* gene content than serotyping (section 3.2). The recovery of identical STs from strains which have been collected from different continents, at different times, reinforces the concept of a clonal population structure for *Bt* (Maynard Smith *et al.,* 1993).

Niche adaptation and plasmid transfer

Bacterial clones are characteristically niche adaptive (Lan & Reeves, 2000; Lan & Reeves, 2001; Cohan, 2002). Some of the clonal lineages recovered from the *B. cereus* group appear to be associated with environmental niche adaptation, e.g. *B. anthracis* in Anthracis, the emetic strains of *B. cereus* in Cereus II and *Bt* israelensis, sotto and other serovars in the Sotto lineage (Figure 3.9). Clonal complexes in *C. jejuni*, linked to environmental niche adaptation, have also been revealed by MLST (Dingle *et al.,* 2002). The separation of *Yersinia pestis* from *Y. pseudotuberculosis* may have been driven by the transfer of a plasmid that encodes virulence factors, allowing the host bacteria to occupy a different ecological niche (Achtman *et al.*, 1999). This may be similar to the situation in the *B. cereus* group where the presence of a plasmid, containing genes

encoding crystal toxins, enables the bacteria to fill the niche of an insect pathogen, or the presence of virulence plasmids allows the bacteria to evolve into a pathogen of humans and / or cattle. Dykhuizen remarked that this is instant speciation (Dykhuizen, 2000).

However, strains from the *B. cereus* group which are associated with wound infections do not belong to a particular lineage but are distributed in clades 1 and 2: - with the exception that no wound pathogens have, so far, been isolated from lineage Sotto or from clade 3 (Barker *et al.,* 2005). This suggests that any bacteria from clade 1 or clade 2 may be opportunistically pathogenic, capable of causing an infection if appropriate conditions arise. However, strains from the Sotto lineage have adapted to fill a specialised niche as successful insect pathogens.

In the *B. cereus* group, as in enteroinvasive *E. coli* (EIEC) / *Shigella,* isolates with the same ST have been recovered from different countries at different times indicating a clonal population. At present, *Shigella* strains are classified in a separate genus from *E. coli*, based on motility and the ability to ferment lactose (EIEC strains are usually negative for these two phenotypic traits). An MLST study of *Shigella* and EIEC isolates showed that most strains of *Shigella* grouped in three clusters within *E. coli.* The authors concluded that EIEC and *Shigella* should not be placed in separate genera, or even separate species but that both should be described as pathovars of *E. coli* (Lan & Reeves, 2001). Strains of *B. cereus* and *B. thuringiensis* cluster together in clade 2 and to a lesser extent in clade 1. *B. anthracis* is a cluster group within the *B. cereus* strains of clade 1. It is logical to deduce that *B. anthracis* and *B. thuringiensis* should be described as pathovars of *B. cereus* and not as separate species. However, such a nomenclature would generate a great deal of confusion.

Genetic exchange in natural populations of soil borne *B. cereus* and *B. thuringiensis* strains has been demonstrated (Vilas-Boas *et al.,* 2002). In this study, two populations were isolated from different locations and examined by MLEE and haemolytic type. Horizontal gene transfer was more prevalent within *B. cereus* or *B. thuringiensis* than between the two species. Although transfer of plasmids between *B. thuringiensis* and *B. cereus* has not been observed in natural situations, indeed it has not been examined, such exchange has been reported under laboratory conditions (Battisti *et al.,* 1985).

However, a NJ tree constructed from the MLEE data recovered two clusters, with strains of *B. thuringiensis* and *B. cereus* represented in each cluster. The authors hypothesised that the lack of genetic differentiation between the two populations may be due to recent divergence with insufficient time for the *B. thuringiensis* and *B. cereus* populations to become completely separated. The authors argued that the species are not genetically isolated from each other and that there is no barrier to horizontal gene transfer, making separate species status difficult to define.

In summary, MLST studies have revealed that for many bacteria inappropriate 'species' status has been given on the basis of serotyping and/or pathogenicity e.g. *Yersinia, E. coli* and *Shigella*. In *S. aureus,* as in the *B. cereus* group, disease causing strains were recovered from different MLST clonal complexes / lineages (Feil *et al.,* 2003, Barker *et al.,* 2005).

Plasmids

Many of the characteristics that are used to differentiate the members of the *B. cereus* group are pathogenicity features that are carried on plasmids. The *B. anthracis* toxin genes are located on pXO1 and the biosynthetic genes for the poly-?-D-glutamic acid capsule are located on pXO2. This is an over-simplification, in that the chromosomal regulatory gene *plcR,* which controls the many virulence genes of *B. cereus* and *B. thuringiensis* is not functional in *B. anthracis* (Agaisse *et al.,* 1999). Because of this, *B. anthracis* is unable to produce many of the extracellular proteins such as phospholipases, enterotoxins and haemolysins which *B. cereus* and *B. thuringiensis* strains manufacture. *B. anthracis* appears to be a very successful clonal group which has recently emerged from a background population of *B. cereus / B. thuringiensis,* by the transfer of plasmids (Read *et al.,* 2003; Rasko *et al.,* 2005). However, the loss of one or both plasmids would make classification of such strains difficult. The discovery of strain G9241, which is a *B. cereus* with the pXO1 virulence genes of *B. anthracis* means that the designation of *B. anthracis* as a separate species from *B. cereus* is even more problematic. Recent discovery of a *B. thuringiensis* strain which produces a poly-?-D-glutamic acid capsule suggests that there may be some plasmid mobility between these groups (Cachat *et al.,* unpublished).

The emetic toxin, cereulide, produced by some *B. cereus* strains is synthesised by a non-ribosomal peptide synthetase genes (Horwood *et al.*, 2004). The emetic toxin producing strains are a clone, ST-26 (Figure 3.9). Recent experiments using plasmid profiling and curing have shown that cereulide production is extrachromosomal and associated with a large plasmid, pCERE01 (Hoton *et al.*, 2005).

B. thuringiensis is characterised and classified according to the insecticidal crystal toxins which, by definition, are synthesised by all strains of *B. thuringiensis*. c*ry* genes are almost invariably located on plasmids. However, in this work, I have uncovered STs that contain strains of both *B. cereus* and *B. thuringiensis*. It may be that *B. thuringiensis* strains loose their plasmids and the ability to kill insects if they adapt to another niche – either in the natural environment or within the constraints of a laboratory environment. Alternatively, these STs may represent recent acquisition of the plasmid by the host and co-adaptation to the new niche has yet to fully emerge. Either way, as plasmids are mobile elements, they are unsuitable criteria for defining a species (Rasko *et al.*, 2005).

The three 'species' *B. mycoides*, *B. pseudomycoides* and *B. weihenstephanensis* have not been reported to have traits for which the genes are located on plasmids. The only distinguishing feature of *B. mycoides* is a peculiar rhizoid growth form on laboratory media such as NA or TSA and this is probably insufficient reason to define *B. mycoides* as a separate species, particularly as the trait can be lost on laboratory subculture (Gordon *et al.*, 1973). *B. pseudomycoides* shares the rhizoid growth habit with *B. mycoides*, but MLST analysis carried out in this study has shown that *B. pseudomycoides* strains are remote from the three clades, although there is evidence for recombination between *B. pseudomycoides* and other strains from the *B. cereus* group, particularly in the *glpF* and *ilvD* alleles. *B. weihenstephanensis* has been awarded separate species status on the grounds of its ability to survive and grow at 4°C but again this is a questionable trait (Stenfors & Granum, 2001).

Proposed Divisions within the *B. cereus* group

Much of the work carried out on bacterial taxonomy has been driven by the study of human diseases and the allocation of species status has often been applied to bacteria because of their particular pathogenic ability or antigenic properties. Unfortunately, the

factors responsible for such traits may not reflect the true phylogeny of the organisms. For example, the acquisition of plasmid encoded virulence genes may transform a benign strain of bacterium into a dangerous pathogen while the genome and evolutionary history (up until the acquisition of the plasmid) remains the same.

The core phylogeny must be determined using robust criteria that reflect, as far as possible, the evolutionary pathways followed. The sequence homology between essential protein coding genes is becoming recognized as a new 'gold standard' for phylogeny.

***B. cereus* group - Phylogenetically based divisions**

Two independent MLST analysis of the *B. cereus* group have been carried out (Helgason *et al.,* 2004) and this one. The study by Helgason and his colleagues involved 77 strains; a large proportion (34), were strains isolated from patients in Norway and the remainder were isolated from soil and dairies. Five strains of *B. anthracis* were included together with various reference strains of *B. cereus* and a *B. weihenstephanensis.* Both MLST studies resolved the data into three groups:- clade 1 (Figure 3.9) corresponds to group III of the study by Helgason *et al*; *B. anthracis* strains, the emetic toxin producers and *B. cereus* ATCC 10987 were recovered in this group. Clade 2 is equivalent to group II of Helgason *et al*; (*B. thuringiensis* and *B. cereus* ATCC 14579) and Clade 3, including *B. weihenstephanensis* and Norwegian soil isolates is similar to group I of Helgason *et al.*

An AFLP study of 154 *B. cereus / B. thuringiensis* soil isolates from Norway, several reference strains and two *B. anthracis* strains allocated the organisms to at least five branches. *B. cereus and B. thuringiensis* strains were mixed together but *B. anthracis* and *B. cereus* ATCC 10987 clustered together (Ticknor *et al.,* 2001). The *B. anthracis / B. cereus* group is again probably equivalent to clade 1 of Figure 3.9 of this thesis. A further, more extensive AFLP analysis (Hill *et al.,* 2004) using strains used in the earlier work by Ticknor *et al.,* (2001) and many additional strains, including 24 *B. anthracis* isolates and over 200 *B. thuringiensis* strains from 36 serovars recovered three main clusters. Clusters 1 and 2 of the AFLP study map almost perfectly to clades 2 and 1, respectively, of this MLST project (strains from cluster 3 in the AFLP study were not examined in this project). The AFLP study recovered two branches, A and C, within cluster 1 which were composed entirely of *B. thuringiensis* strains. Branch A

corresponds to lineage Sotto in this thesis (Figure 3.9), being composed of isolates from *B. thuringiensis* serovars darmstadiensis, israelensis morrisoni and sotto which are all successful insect pathogens. Branch C corresponds to lineage Kurstaki, with strains from serovars aizawai and galleriae included in both studies. Other branches in clusters 2 and 3 contained mixtures of *B. thuringiensis* and *B. cereus*. The *B. anthracis* isolates were monomorphic in the AFLP analysis and were recovered in branch F of cluster group 2. Pathogenic strains of *B. cereus* were widely dispersed on the AFLP tree, but 22 out of 42 isolates were recovered in cluster group 2 and 16 of those were found to be on branch F, close to *B. anthracis*. *B. cereus* strains isolated from wound pathogens were recovered across clades 1 and 2 (Figure 3.62). Such good agreement between the AFLP study and the results of my MLST analysis gives strong support for the validity of the clade assignments found in this work and justifies the proposal of a new classification system for the *B. cereus* group.

For strains of the *B. cereus* group examined in this thesis, a NJ tree constructed from concatenated sequence data from seven housekeeping genes separated the 146 isolates (88 STs) into three clades, which further sub-divided into nine distinct groups or lineages. Statistical techniques using pairwise comparisons have been applied to compare 'Fst', 'Fixed Differences' and 'Shared Mutations' for the six species of the *B. cereus* group and found low values for Fst, no fixed differences and 246 shared mutations between *B. cereus* and *B. thuringiensis* (Section 3.1.4). When the statistical tests were used to examine the data split into lineages, generally high values for Fst with many fixed differences and few shared mutations confirming the validity of assigning lineages.

Proposed new classification of the *B. cereus* group – 3 clades and 9 lineages.

The NJ tree of concatenated sequences was used to group the STs into 3 clades and 9 lineages (Figure 3.9).

Clade 1 was composed of 24 STs, *B. anthracis*, *B. cereus* and two rare *B. thuringiensis* strains. Clade 1 clustered into a clone of *B. anthracis* strains and three Cereus lineages :- Cereus I, Cereus II and Cereus III. There were some unassigned strains and it is likely that as more strains are added to the database, more distinct lineages will arise.

Clade 2 was made up of 55 STs, predominantly *B. thuringiensis* (32 STs), but with many *B. cereus* strains (25 STs) represented. There were two cases where strains of *B. cereus* and *B. thuringiensis* shared the same ST. Clade 2 clustered into four lineages :- Sotto, Thuringiensis, Tolworthi and Kurstaki, named to reflect the position of the major existing serovars of *B. thuringiensis.*

Clade 3 contained *B. mycoides, B. weihenstephanensis* and some *B. cereus* isolates from Norwegian soil. This is a heterogeneous clade, which I have referred to as 'the others' and more work needs to be carried out to determine the structure of this group. The *B. weihenstephanensis* strains and the *B. cereus* strains from Norway do not share the rhizoid growth habit of *B. mycoides.*

A small fourth group was identified as the 'outliers', and contained only two strains: - a *B. pseudomycoides,* and a strain that was received as *B. mycoides,* but was probably mis-identified, and is a strain of *B. pseudomycoides*. The outliers were so far removed from the other clades, that *B. pseudomycoides* was used to provide a root for the phylogenetic tree shown in Figure 3.9. This group could retain separate species status or be known as lineage Pseudomycoides, but more strains would need to be examined to confirm the status of this group.

Lineages within clade I

Clade 1, Lineage Anthracis

The Anthracis lineage was found to be very strongly clonal. Only three nucleotide changes were identified among the 11 strains examined, but these enabled separation into three groups. Eight strains were ST-1 and identical to the Ames strain which was isolated from a cow in Texas in 1981 (Read *et al.,* 2004). One strain, the US vaccine strain, ST-2 (V770-NPI-R) was found to have single base changes in two alleles (*ilvD* and *pycA*) (Figure 3.10). Two strains belonging to the 'B' group of *B. anthracis* (ST-3) were separated from the more common 'A' group on the basis of a single nucleotide change in *glpF*. The 'B' group has been further divided into B1 and B2 by multiple-locus variable-number tandem repeat analysis (Keim *et al.,* 2000). The B1 genotypes (represented by strain K2478 from Mozambique) are geographically restricted to southern Africa and B2 genotypes (represented by strain K2762 from France) are

exclusive to Europe. This finding of a high degree of clonality in *B. anthracis* is consistent with the published data (Keim *et al.,* 2000, Hill *et al.,* 2004, Rasko *et al.,* 2005).

The non-haemolytic enterotoxin gene *nhe* was detected in *B. anthracis* by a BLAST search using *B. cereus nhe* sequence data against the *B. anthracis* sequence at TIGR. Interestingly, when a search was carried out using the *hbl* sequence, the nearest hit was again *nhe,* indicating that *hbl* is absent. Some similarities between the amino acid sequences from Hbl (L_1) and the 39-kDa protein from Nhe have been reported (Lund & Granum, 1997). However, in *B. anthracis,* the regulatory gene *plcR*, which controls *nhe* and *hbl* is not active due to mutation (Read *et al.,* 2003).

Clade 1, lineage Cereus I

Cereus I was composed of four strains of *B. cereus* and included ATCC 10987 which is an atypical xylose-positive strain isolated in Canada in 1930 from spoiled cheese. Strain *B. cereus* M21was isolated from a cheese dairy in Finland in 1998. Strain *B. cereus* m1545 was isolated from powdered milk in Brazil (1987) and strain *B. cereus* m1564 was recovered from dried coconut in Brazil (1987). While this lineage could not be described as clonal, three out of four strains were associated with milk products and this may suggest a common ancestry. The genome of *B. cereus* ATCC 10987 has been sequenced and was found to be much closer to *B. anthracis* than to the *B. cereus* type strain (ATCC 14579^T) (Rasko *et al.,* 2004).

Clade 1, lineage Cereus II

Cereus II contained four STs and all were *B. cereus.* ST-26 comprised a clonal group of five strains – three were recovered from emetic food poisoning outbreaks in the UK (F3080 B, F3942/87, F4810/72) and strain F4746 was isolated from fried rice. The remaining isolate (S710) was isolated from soil and was found to synthesize cereulide (1.6 ng/ml of culture) (Priest *et al.,* 2004). The other STs in Cereus II included ST-31, *B. cereus* S366 isolated from the North Sea, ST-45, *B. cereus* m1293 from cream cheese (Brazil, 1987) and ST-47, *B. cereus* m1576 from uncooked chicken (Brazil, 1987) The remaining strain in this lineage was ST-81, *B. cereus* S78, (ATCC 7004).

None of the emetic toxin-producing strains of *B. cereus* were able to hydrolyse starch, a feature that has been reported to be typical of these strains (Ehling-Schulz *et al.*, 2005). Additionally, they did not have the *hbl* gene (Figure 3.60 and Table 3.30). Strain m1293 (ST-45) strain m1576 (ST-47) and S78 (ST-81) were also unable to hydrolyse starch, making this feature common to all strains of lineage Cereus II. Strain m1293 did not have *nhe* or *hbl*, S366 and S78 had both *nh*e and *hbl* genes (Figure 3.62).

Clade 1, Lineage Cereus III

There were eight STs in lineage Cereus III, which was the closest lineage to Anthracis.

Three of the STs were associated with human infections. ST-75 (*B. cereus* R3039/03) was isolated from septicaemia and had two alleles (*gmk-1* and *pta-1*) in common with *B. anthracis*. ST-76 (*B. cereus* 191560K) was isolated from an infected insect bite and ST-77 (*B. cereus* 168287M) was isolated from a haematoma from a patient following an operation. Two STs, ST-11 and ST-62, were *B. cereus* strains deposited in the database by the Centers for Disease Control and Prevention, Atlanta, USA, and probably represented pathogens. ST-27 (*B. cereus* F4370/75) was isolated from barbecued chicken and had one allele (*gmk-1*) in common with *B. anthracis.* The two remaining STs were found to be unusual strains of *B. thuringiensis.* ST-57 was a strain of *Bt* darmstadiensis, from Pakistan and ST-60 was a strain of *Bt* kumamotoensis from Iraq. These two *B. thuringiensis* strains were atypical of their serovars with unique STs and RAPD patterns. There was some evidence that recombination involving the *ilvD* allele had occurred, as there were 24 nucleotide changes between ST-57 and ST-27 (Figure 3.13).

The enterotoxin genes *nhe* and *hbl* were not present in ST-60, ST-76 or ST-77. The *hb*l gene was absent from ST-57 while *nhe* was present. However, ST-27 and ST-75 possessed both enterotoxin genes (Figure 3.62).

Unassigned STs of clade 1

Some STs within clade 1 were not assigned to a lineage. Two of these STs were close together - *B. cereus* S360 (ST-30), isolated from the North Sea and *B. cereus* K8 (ST-82) from a cheese dairy. Two other unassigned strains were *B. cereus* R3098/03 (ST-74)

isolated from septicaemia and *B. cereus* G9241 (ST-78) which was recovered from a patient with a disease similar to inhalation anthrax. (Hoffmaster *et al.*, 2004).

Overall composition of clade 1

Clade 1 was composed mainly of strains designated as *B. cereus*, but two rare strains of *B. thuringiensis* were also recovered in clade 1. *B. anthracis* is a clonal group found within clade 1. Enterotoxin genes were less common in clade 1 than clade 2. Of the STs tested, *nheBC* was found in 64 % of STS and *hblCD* was recovered from 40% of STs (Table 3.29).

Lineages within clade 2

Clade 2, lineage Sotto

Lineage Sotto was composed entirely of *B. thuringiensis* strains and represents a clonal complex. ST-12, ST-16 and ST-23 had five alleles in common - *glpF, gmk, ilvD, pta* and *tpi* (Figure 3.14). It was interesting to note that the three strains within ST-16 were different serotypes. The ST-16 *B. thuringiensis* strains were israelensis CCCT-2259, morrisoni T08025 and canadensis T05a030. All three ST-16 strains had identical RAPD patterns. This RAPD pattern was found in all but one strain of the *B. thuringiensis* serovar israelensis strains examined by RAPD (Figure 3.42). The israelensis strain with an unusual RAPD pattern (an extra band), T14531, was isolated from Vietnam in 1998. MLST analysis revealed that strain T14531 (ST-65) had two different alleles (*ilvD-14* and *pycA-14*) from israelensis CCCT 2259 (ST-16).

ST-23 was represented by five strains of *B. thuringiensis* serovar morrisoni and one strain of *B. thuringiensis* pathovar tenebrionis. RAPD patterns for the five strains of *Bt* morrisoni and *Bt* tenebrionis were the same. However, there was a difference which was not detected by MLST or RAPD – the morrisoni serovars contained *cry1* toxin genes (Lepidopteran specific) whereas the tenebrionis strain contained cry*3a* toxin genes (Coleopteran specific) (Table 3.25).

ST-12 was made up of three strains of *Bt* sotto and one strain of *Bt* dakota. ST-56 comprised four strains of *Bt* darmstadiensis, with similar RAPD patterns and ST-61 was

represented by a single strain of *Bt* galleriae. The remaining STs in the lineage Sotto occurred only once each and all were serovars of *B. thuringiensis*: - ST-49, sotto; ST-55, darmstadiensis; ST-65, israelensis; ST-67, morrisoni; ST-69, morrisoni and ST-70, morrisoni. Recombination in lineage Sotto was indicated from the >26 nucleotide changes in the *ilvD* gene, splitting the lineage as illustrated in the SplitsTree graph (Figure 3.14).

The enterotoxin genes *nheBC* and *hblCD* were present in most strains from all STs of the Sotto lineage (*nhe* was not recovered from ST-69 and *hbl* was not found in ST-61).

The Sotto lineage of clade 2 is highly clonal and composed entirely of closely related, apparently successful insect pathogens. The presence of enterotoxin genes is common in this lineage.

Clade 2, lineage Thuringiensis

Lineage Thuringiensis is mixed, containing two STs of *B. cereus* strains and four STs of *B. thuringiensis* strains. ST-10 is made up of five strains of *Bt* serovar thuringiensis, T01001, T01015, T01022, T01058 and T01326; all had identical RAPD patterns. Figure 3.41 shows a high degree of homogeneity in the 20 strains of *Bt* thuringiensis examined by RAPD: - 19 were identical. ST-10 represents a clonal group of serovar thuringiensis strains containing *cry1 / 1b* toxin genes (Lepidopteran specific). Interestingly, one strain of thuringiensis, T01058, with ST-10, RAPD pattern 1 and *cry1* toxin gene was found to also have the *cry3a* gene, which makes it potentially pathogenic to Coleopteran insects as well as Lepidoptera. The presence of the additional *cry* gene was not detected by RAPD. Both enterotoxin genes *nheBC* and *hblCD* were present in all STs of lineage Thuringiensis, although not recovered from all representatives of each ST.

Bt thuringiensis is a clonal group but the other STs present in lineage Thuringiensis only have one or two alleles in common with ST-10 (Figure 3.15, Table 3.1 and Table 3.15). The STs in this category are represented only once in the dataset - ST-20 (*B. cereus* WSBC 10312), ST-43 (*B. cereus* m1278), ST-58 (*Bt* dakota T15006), ST-63 (*Bt* morrisoni T08019) and ST-71 (*Bt* tolworthi T09020).

Clade 2, lineage Tolworthi

Lineage Tolworthi is a large, mixed group comprising 16 STs (eight *B. cereus* and eight *B. thuringiensis*). Nevertheless, there is evidence for the emergence of clonal complexes as many of the STs were represented by several strains. ST-4 represented by the type strain of *B. cereus* (ATCC 14579^T) was in this Lineage. ST-14 comprised two strains of *Bt* entomocidus; ST-17 was composed of two strains of *Bt* pakistani; ST-19 was made up of two strains of *B. cereus*, WSBC 10249 (isolated from soil) and m1280 (isolated from vegetable soup); ST-22 comprised five strains of *Bt* tolworthi; ST-24 contained three strains of *B. cereus*, NCIB 6349, cal3 (isolated from food packaging board machinery) and WSBC 10028 (isolated from pasteurised milk) and ST-73 represented two *B. cereus* strains isolated from cases of septicaemia. The other STs only occurred once. They were ST-29 (*B. cereus* S86), ST-34 (*B. cereus* ATCC 11778), ST-46 (*B. cereus* m1550, isolated from uncooked chicken), ST-48 (*Bt* thuringiensis T01246, an atypical thuringiensis strain, isolated from Iraq), ST-50 (T05a001) and ST-52 (T05a019) (both strains of *Bt* canadensis), ST-53 (*Bt* aizawai T07146), ST-64 (*Bt* morrisoni T08003) and ST-86 (*B. cereus* TSP10 isolated from food packaging board).

The Tolworthi lineage was heterogeneous, with possible clonal groups emerging. The SplitsTree graph (Figure 3.16) indicated that lineage Tolworthi is developing into two sub-lineages, and the division appears to correlate with a recombination event in the *ilvD* allele. Taking ST-22, in the sub-lineage to the left as the reference point, all the STs with >14 nucleotide changes in *ilvD* were found in the sub-lineage to the right. The two enterotoxin genes were present in most STs, but the products could not be amplified from DNA of *B. cereus* ATCC 11778, (ST-34) and DNA from *Bt* morrisoni T08003, (ST-64) did not act as a template for the *hbl* gene.

The *cry* gene content of strains within this lineage included *cry1e* (Figure 3.55 and Table 3.26) from ST-22 strains (5 representatives of *Bt* tolworthi). ST-48 (an atypical *Bt* thuringiensis) did not give a product with *cry1 or cry3a* primers and ST- 63 and ST-64 (unusual *Bt* morrisoni strains) failed to amplify a product with *cry1*, *cry3a* or *cry4* primers. These atypical STs consistently failed to amplify *cry* genes by PCR (Tables 3.25 and 3.26). Other *B. thuringiensis* serotypes were also represented in lineage

Tolworthi and included aizawai (ST-53), canadensis (ST-50 and ST-52), entomocidus (ST-14) and pakistani (ST-17).

Clade 2, lineage Kurstaki

The Kurstaki lineage, like Tolworthi, was mixed, with 10 *B. cereus* STs and nine *B. thuringiensis* STs. This lineage was the only one to contain STs which represented strains of both *B. cereus* and *B. thuringiensis*. This occurred twice: - ST-18 contained *Bt* pakistani T13028 and *B. cereus* 172560W (isolated from a burn to the leg). ST-8 comprised five strains of *Bt* kurstaki and three strains of *B. cereus*, which had been isolated from mud in Maryland. Of these, *B. cereus* S58 was associated with swan faeces. ST-13 was a clonal group composed of five strains of *Bt* kenyae; ST-15 was represented by three strains of *Bt* aizawai; ST-25 comprised two strains of *Bt* pakistani; ST-59 two strains of *Bt* kumamotoensis and ST-68 two strains of *Bt* tolworthi. The remaining STs were represented once only: - ST-33, *B. cereus* ATCC 10876; ST-39, *B. cereus* SPS2 (from mineral pigment); ST-40, *B. cereus* TSP11 (from food packaging board); ST-44, *B. cereus* m1292 (from chocolate UHT milk); ST-51, *Bt* canadensis T05a015; ST-54, *Bt* aizawai; ST-66, *B. cereus* m1552 (from boiled potato); ST-84 and ST-85 were *B. cereus* strains submitted to the MLST database by the Centers for Disease Control and Prevention, Atlanta, USA.

One strain, *B. cereus* TSP8 (ST-80), isolated from food packaging did not possess *nheBC* and *hblCD*. The two enterotoxin genes were detected from some strains within each ST, but were not amplified from all representatives of each ST. Overall, the presence of enterotoxin genes was common in strains of the Kurstaki lineage.

An examination of the ST-8 strains of *B. cereus* (S57, S58, S59) and *Bt* kurstaki (T03a001) by SDS-page recovered Cry1 protein from *Bt* kurstaki (T03a001) and a Western blot detected a small amount of Cry1 protein from *B. cereus* S58 (Figure 3.56). This may suggest that that the *B. cereus* strains were originally *B. thuringiensis* but had lost their plasmids, perhaps because they have been confined to a laboratory environment for about 30 years.

Numerous *B. thuringiensis* serotypes other than *Bt* kurstaki were recovered within the Kurstaki lineage - ST-15 and ST-54 (aizawai), ST-51 (canadensis), ST-25 (galleriae),

ST-13 (kenyae), ST-59 (kumamotoensis), ST-18 (pakistani) and ST-68 (tolworthi). The two *Bt* tolworthi strains were both assigned to ST-68 (T09016 and T09043) and DNA from these organisms did not provide template for *cry1* amplifications. These strains did not share any alleles with the ST-22 strains from the Tolworthi lineage (Table 3.26). The Kurstaki lineage was complex and the SplitsTree graph (Figure 3.17) showed that it could be divided into three sub-lineages.

Unassigned STs of clade 2

Six *B. cereus* STs were unassigned in clade 2. They were ST-9, NCTC 6474; ST-28, F4431/3 (from rice); ST-38, ATCC 4342; ST-79, S79 and ST-80, TSP8 (isolated from food packaging board).

Overall composition of clade 2

The overall composition of clade 2 was 42% *B. cereus* and 58% *B. thuringiensis* strains. Only one lineage (Sotto) consisted exclusively of *B. thuringiensis* strains. Probable recombination events involving the *ilvD* gene were detected in lineages Sotto and Tolworthi. Recovery of about equal numbers of strains of *B. cereus* and *B. thuringiensis* in the lineages Tolworthi and Kurstaki provided strong evidence that *B. cereus* and *B. thuringiensis* may not reasonably be described as separate species. This was reinforced by two instances in lineage Kurstaki where strains of *B. cereus* and *B. thuringiensis* had identical STs and by definition were clonal. The enterotoxin genes were very common in clade 2. Of the STs examined, 96% had *nheBC* in some or all strains of the ST and *hblCD* was indicated in 88 % of STs.

Clade 3

Clade 3 was a mixed group, which I referred to as the ‘others’. Seven STs were found, each with only a single strain. The STs were ST-21 *B. mycoides* WSBC 10277 (from pasteurised milk). STs 35, 36 and 37 represented *B. cereus* strains isolated from soil in Norway, ST-41 was *B. weihenstephanensis* WSBC 10202 (from pasteurised milk) and ST-42 represented *B. weihenstephanensis* WSBC 10364 (from soil) and ST-88 was *B. mycoides* STR4 (from soil in Scotland). As there are >30 nucleotide changes, recombination is likely to have occurred in the *ilvD* gene for ST-36 and ST-37

(Norwegian soil isolates) and ST- 42 (*B. weihenstephanensis*) and in *gmk* for ST-21 (*B. mycoides)* (Figure 3.18).

The *B. mycoides* strains were characterised by a rhizoid form of growth on laboratory media such as NA or TSA. Five out of the seven strains in this clade had the enterotoxin gene *nheBC*, but only one strain (STR4) was positive for *hblCD*.

The 'outliers'

These two STs were so far removed from the other clades that I used one of them (ST-83) as a root for the phylogenetic tree shown in Figure 3.9. The STs in this group were ST-83, *B. pseudomycoides* WS 3118 (from soil in the USA) and ST-87, a strain received as *B. mycoides*, but probably mis-identified and correctly placed as *B. pseudomycoides*.

Strains of *B. pseudomycoides* are distant from the strains of clades 1, 2 and 3 and perhaps *B. pseudomycoides* should retain separate species status. This is in accordance with Nakamura (1998) who showed that *B. mycoides* strains segregated into two groups, *B. mycoides* sensu stricto and *B. pseudomycoides*, with DNA relatedness of about 30% and had different fatty acid compositions. Relatedness within each group was 70% to 100%. More strains from these species need to be examined, especially as there is evidence for recombination between *B. pseudomycoides* and other strains of the *B. cereus* group (Section 3.1.6).

Conclusions

From the MLST data it can be deduced that there is no basis for considering *B. cereus* and *B. thuringiensis* to be separate species. Furthermore, using MLST data, the clade 3 strains, *B. weihenstephanensis* and *B. mycoides* do not appear to be sufficiently different from *B. cereus*, to warrant species status. *B. pseudomycoides*, the outlier group, is further removed from the three clades and probably represents a separate species. *B. anthracis* appears to be a homogeneous group within clade 1 and because of its pathogenic nature it is pragmatic to retain species status.

The present nomenclature is confusing. One approach would be to return to a system not unlike that favoured by Gordon *et al.* (1973) which adhered to an earlier nomenclature adopted by Smith *et al.* (1946), in which all strains of the *B. cereus* group were considered to be members of the species *B. cereus* with varietal names used for *B. anthracis, B. mycoides* and *B. thuringiensis*. In this phylogenetically based version of the scheme, while species names are retained to label individual strains, a lineage system could be adopted which allocates STs to phylogenetic groups.

Clade 1 Predominantly *B. cereus*

Lineage Anthracis

Lineage Cereus I

Lineage Cereus II (emetic group)

Lineage Cereus III

Clade 2 Predominantly *B. thuringiensis*

Lineage Sotto

Lineage Thuringiensis

Lineage Tolworthi

Lineage Kurstaki

A strain of *B. thuringiensis* could be defined as a Cry positive *B. cereus.*

Clade 3 Predominantly *B. mycoides*

Although too few strains have been studied for far-reaching conclusions, indications are that *B. mycoides* will predominate here, i.e. All strains of *B. mycoides* have been allocated to Clade 3.

Lineage Mycoides (includes *B. mycoides, B. weihenstephanesis* and other strains which are recovered in clade 3)

Lineage *B. pseudomycoides*

The Pseudomycoides Lineages may be a separate species, but more strains need to be examined to confirm this.

International Database for the Multi-Locus Sequence Typing of *Bacillus*.
http://pubmlst.org/bcereus

Notwithstanding the recent expansion of whole genome sequencing, the sequencing of seven housekeeping genes will be simpler and less expensive for the next five years or so and the MLST scheme should provide a valuable database. The database is steadily expanding. Since completion of the data collection for this thesis, the number of strains deposited in the database has risen from 146 to 235 and the number of STs recovered has increased from 88 to 149 (August 2005). Contributions to the database are received from several research groups within the Centers for Disease Control and Prevention and from scientists in S. Korea and Japan. STs and strains from the database have been cited in recent publications (Hoffmaster *et al.*, 2004, Kim *et al.*, 2004, Priest *et al.*, 2004, Rasko *et al.*, 2005, Barker *et al.*, 2005).

A continually expanding collection of the sequence types of strains from the *B. cereus* group is of great benefit to the whole community. The database is freely available to all and can be analyzed by a variety of phylogenetic techniques and evolutionary models constructed. The MLST database has already shown that *B. cereus* strains isolated from clinical infections are genomically diverse and not restricted to clonal complexes (Barker *et al.*, 2005).

The work carried out for this Ph. D. thesis will, I hope, form the beginning of a much greater project.

APPENDIX A

Strains of *Bacillus anthracis* (DNA only), *Bacillus cereus, Bacillus mycoides, Bacillus thuringiensis* and *Bacillus weihenstephanensis* used in this project.

Source codes :-

1 *Bacillus cereus* strains from the collection of Professor F.G. Priest at Heriot-Watt University.

2 *Bacillus cereus / Bacillus thuringiensis* strains from the collection of Anne-Brit Kolstø, Biotechnology Centre of Oslo and Institute of Pharmacy, University of Oslo, Norway.

3 *Bacillus cereus, Bacillus mycoides, Bacillus pseudomycoides, Bacillus thuringiensis* and *Bacillus weihenstephanensis* strains from the collection of Siegfried Scherer., Institut fur Mikrobiologie, FML Weihenstephan, Technische Universitat Munchen.

4 *Bacillus cereus* strains from the collection of Mirja S. Salkinoja-Salonen, Department of Applied Chemistry and Microbiology, University of Helsinki, Finland.

5 *Bacillus cereus* strains from Brazil. Collection of Marilena d'Aquino de Muro (former PhD student and Post-doc, Heriot- Watt University). 1987

6 *Bacillus thuringiensis* strains from the Institut Pasteur, Paris. (begin with 'T').

7 *Bacillus anthracis* strains, provided by Dr L. Baillie (DNA only)

8 *Bacillus cereus* strains, wound pathogens. Provided by Dr B.Thakker at the Western Infirmary, Glasgow.

9 *Bacillus cereus* strains from J. McLauchlin, Health Protection Agency, London. ('R' numbers).

10 Additional strains of *Bacillus thuringiensis*, aquired towards the end of the project, for further investigation into the structure of selected serotypes, from the Institut Pasteur (T strains)

11 *Bacillus thuringiensis* strains purchased from the Bacillus Genetic Stock Centre (BGSC).

Table 5.1 List of Strains

Strain	Species	Sero type	Other numbers / details	Country	Date	Source
S 57	*cereus*		B. Austin SA 215	Maryland USA	1975	1
S 58	*cereus*		B. Austin SA 219 swan faeces.	Maryland USA	1975	1
S 59	*cereus*		B. Austin SA 224	Maryland USA	1975	1
S 61	*cereus*		P.A. Hartman, 423, IA 27.			1
S 62	*cereus*		P.A. Hartman, 424, IA 36. Y.L. Quinn, ATCC 11778.			1
S 63	*cereus*		P. A.Hartman, 459, NRRL B-344.			1
S 64	*cereus*		NCIB 6349			1
S 66	*cereus*		NCTC 6474 T.R.G. Gray, B20.	UK	1988	1
S 67	*cereus*		T. R. G. Gray, B 21. NCTC 7464			1
S 68	*cereus*	5	R. J. Gilbert, 3502 / 73, fried rice.		1973	1
S 70	*cereus*	6	R. J. Gilbert, 4370 / 75, barbecued chicken.	UK	1975	1
S 71	*cereus*	8	R. J. Gilbert, 4431 / 73, Indonesian rice dish.	Indonesia	1973	1
S 72	*cereus*		R. J. Gilbert, 4433 / 73, meat loaf.		1973	1
S 73	*cereus*	1	R. J. Gilbert, 4810 / 72, vomit.		1972	1
S 74	*cereus*	1	R. J. Gilbert, 4746 / 77. fried rice		1977	1
S 75	*cereus*		H. J. Somerville, T1 87 strain T			1
S 76	*cereus*		P. A. Hartman, 461. NRS 996.			1
S 77	*albolactis*		NCIB 5097			1
S 78	*albolactis*		NCIB 8079			1
S 79	*fluorescens*		H. J. Somerville, T1 53. NCIB 2600. 300			1
S 86	*terminalis*		WR 7100.		1988	1
S 363	*cereus*		G. J. Bonde, 354 Marine	N. Sea		1
S 366	*cereus*		G. J. Bonde, 372 Marine	N. Sea		1
S 710	*cereus*		5893			1
DSM 31	*cereus*		J. R. Norris, BO 002. ATCC 14579			1
AH 75	*cereus*		ATCC 10987 Milk / Milk products	Canada	1930	2
AH 181	*cereus*		ATCC 10876			2
AH 182	*cereus*		ATCC 11778			2
AH 517	*cereus / thuring.*	H 35	Crystal inclusion attached to spore	Norway, Moss 1	1998	2
AH 563	*cereus / thuring.*	H -	Crystal inclusion attached to spore	Norway, Moss 4.	1998	2

Strain	Species	Sero type	Other numbers / details	Country	Date	Source
AH 581	*cereus / thuring.*	H 24	Crystal inclusion attached to spore	Norway, Moss 4.	1998	2
AH 621	*cereus / thuring.*	H 34		Norway, Singsas.	1998	2
AH 636	*cereus / thuring.*	H 11	Toxic against Trichoplusia ni larvae	Norway, Singsas.	1998	2
AH 647	*cereus / thuring.*	H 3a, 3b,3c		Norway, Steigen.	1998	2
AH 651	*cereus / thuring.*	H 34		Norway, Steigen.	1998	2
AH 656	*cereus / thuring.*	H 34		Norway, Brekkvos selv	1998	2
AH 662	*cereus / thuring.*	H 3a, 3d,3e		Norway, Brekkvos selv.	1998	2
AH 671	*cereus / thuring.*	H 34		Norway, Tromsø	1998	2
AH 681	*cereus / thuring.*	H 3a, 3b, 3c	Toxic against Trichoplusia ni larvae	Norway, Tromsø.	1998	2
AH 684	*cereus / thuring.*	H 3a, 3b, 3c		Norway, Tromsø.	1998	2
AH 685	*cereus / thuring.*	H 3a, 3b, 3c		Norway, Tromsø.	1998	2
AH 686	*cereus / thuring.*	H 3a, 3b, 3c	Toxic against Trichoplusia ni larvae	Norway, Tromsø	1998	2
AH 691	*cereus / thuring.*	H 3a, 3b, 3c		Norway, Tromsø	1998	2
AH 693	*cereus / thuring.*	H 3a, 3b, 3c		Norway, Tromsø	1998	2
AH 821	*cereus*		ATCC 4342	USA	1900	2
AH 923	*cereus*		ATCC 14579, DSM 31	USA	1916	2
WSBC 10028	*cereus*		Pasteurised milk	Germany	1999	3
WSBC 10249	*cereus*		Soil, (S 38, 24)	Denmark	1999	3
WSBC 10312	*cereus*		Kurkuma root	Thailand	1999	3
WSBC 10360	*mycoides*		Soil.	Germany	1999	3
WSBC 10277	*mycoides*		Pasteurised milk	Germany	1999	3
WS 3118	*pseudo mycoides*		Soil, NRRL B - 617	USA.	1999	3
WS 2629	*thuringiensis*		HER 1418		1999	3
WS 2734	*thuringiensis*		DSM 2046		1999	3
WSBC 28002	*thuringiensis*		Soil, (P02)	Poland	1999	3
WSBC 10202	*weihen stephanensis*		WS 2478. Pasteurised milk	Germany	1999	3
WSBC 10204	*weihen stephanensis*		WS 2480. Pasteurised. milk	Germany	1999	3
WSBC 10364	*weihen stephanensis*		Soil.	Germany	1999	3

Strain	Species	Sero type	Other numbers / details	Country	Date	Source
F4810 /72	*cereus*		Emetic toxin. SMR 178 Emetic outbreak	USA	1972	4
F3942 /87	*cereus*		Emetic toxin. Boiled rice Emetic outbreak	UK	1987	4
F3080B /87	*cereus*		Emetic toxin. Chicken Korma/rice. Emetic outbreak	UK	1987	4
M21	*cereus.*		Cheese dairy	Finland	1998	4
H 5	*cereus*		Cheese dairy			4
TSP11	*cereus.*		Food packaging board		1999	4
BM-VIIIP1	*cereus*		Board machine, slime.			4
Cal3	*cereus.*		Board machine, calender water	Finland		4
TSP10	*cereus*		Food packaging board			4
SPS 2	*cereus*		Mineral pigment Kaolin		1999	4
TSP8	*cereus.*		Food packaging board			4
K8	*cereus*		Cheese dairy	Finland		4
m1282	*cereus*		Vanilla profiterole (1.85×10^4) 90/1, 90/4	Brazil	1987	5
m1291	*cereus.*		Birthday cake (4.5×10^7g) 280/1	Brazil	1987	5
m1550	*cereus*		Uncooked chicken (1.3×10^2g) 277/2	Brazil	1987	5
m1558	*cereus*		Polenta 88/1	Brazil	1987	5
m1552	*cereus.*		Boiled potatoes (5.0×10^3g) 166/4	Brazil	1987	5
m1576	*cereus*		Uncooked chicken (-) 82/2	Brazil	1987	5
m1584	*cereus*		Boiled rice (-) 84/2	Brazil	1987	5
m1592	*cereus.*		Savoury crepe (-) 86/1	Brazil	1987	5
m90 / 4	*cereus*		Vanilla profiterole (1.85×10^4)	Brazil	1987	5
m218 / 3	*cereus*		Frozen tropical fruit pulp (1.1×10^3 g)	Brazil	1987	5
m1557	*cereus*		Cassava flour (6.0×10^2g) 47/2	Brazil	1987	5
m213 / 5			Complete food for intubate feeding . (>1.0×10^3 ml)	Brazil	1987	5
m166 / 5			Boiled potatoes (166 / 15.0×10^3g) 166 / 1, 1594	Brazil	1987	5
m1278			Strawberry ice cream ($2x10^2$/g) 4/1	Brazil	1987	5
m1280			Vegetable soup (>5 $x10^3$/g) 71/2	Brazil	1987	5
m1292			Chocolate UHT milk (-) BN-1	Brazil	1987	5
m1293			Cream cheese (-) BPS-2	Brazil	1987	5
m1545			Full fat powdered milk ($5x10^2$/g) 77/1	Brazil	1987	5
m1564			Dried grated coconut ($2x10^2$/g) 204/2	Brazil	1987	5
m1568			Vanilla custard ($2x10^2$/g) 266/1	Brazil	1987	5

Strain	Species	Sero type	Other numbers / details	Country	Date	Source
m1574			Reconstituted milk 2% ($2x10^2$/ml) 36/2	Brazil	1987	5
m95/5			Frozen cooked king prawns (>$5x10^3$/g)	Brazil	1987	5
m11/4			Mussarella cheese (>$5x10^3$/g)	Brazil	1987	5
m96/1			Uncooked hot dog sausage (>$5x10^3$/g)	Brazil	1987	5
m92/5			Cooked courgette + cassava flour ($3x10^3$/g)	Brazil	1987	5
m221/2			Pizza ($6x10^2$/g)	Brazil		5
T01001	*thuringiensis*	1		Canada	1958	6
T01015	*thuringiensis*	1		Bulgaria	1962	6
T01022	*thuringiensis*	1		U.S.A.	1964	6
T01246	*thuringiensis*	1		Iraq	1984	6
T01326	*thuringiensis*	1		Chile	1994	6
T03a001	*kurstaki*	3a, 3b, 3c		France	1962	6
T03a075	*kurstaki*	3a, 3b, 3c		Iraq	1976	6
T03a172	*kurstaki*	3a, 3b, 3c		Pakistan	1982	6
T03a287	*kurstaki*	3a, 3b, 3c		Kenya	1988	6
T03a361	*kurstaki*	3a, 3b, 3c		Australia	1990	6
T04002	*sotto*	4a, 4b		Canada	1965	6
T04016	*sotto*	4a, 4b		Pakistan	1980	6
T04024	*sotto*	4a, 4b		Pakistan	1981	6
T04236	*sotto*	4a, 4b		Indonesia	1991	6
T04b001	*kenyae*	4a, 4c		Kenya	1962	6
T04b012	*kenyae*	4a, 4c		Bulgaria	1974	6
T04b054	*kenyae*	4a, 4c		Iraq	1986	6
T04b060	*kenyae*	4a, 4c		Iraq	1987	6
T04b073	*kenyae*	4a, 4c		Chile	1993	6
T05005	*galleriae*	5a, 5b		U.S.A.	1964	6
T05033	*galleriae*	5a, 5b		U.S.A.	1975	6
T05114	*galleriae*	5a, 5b		France	1985	6
T05125	*galleriae*	5a, 5b		South Korea	1993	6
T05a001	*canadensis*	5a, 5c		Canada	1968	6
T05a015	*canadensis*	5a, 5c		U.S.A.	1977	6
T05a019	*canadensis*	5a, 5c		Pakistan	1980	6
T05a030	*canadensis*	5a, 5c		Iraq	1987	6
T06007	*entomocidus*	6		Pakistan	1983	6
T06010	*entomocidus*	6		Pakistan	1983	6
T07033	*aizawai*	7		Japan	1975	6
T07058	*aizawai*	7		France	1983	6
T07146	*aizawai*	7		Indonesia	1991	6
T07180	*aizawai*	7		Spain	1992	6
T07196	*aizawai*	7		Brazil	1993	6
T08001	*morrisoni*	8a, 8b		U.S.A.	1963	6
T08009	*morrisoni*	8a, 8b		U.S.A.	1979	6
T08012	*morrisoni*	8a, 8b		Pakistan	1980	6
T08023	*morrisoni*	8a, 8b		Brazil	1987	6
T08025	*morrisoni*	8a, 8b		France	1988	6

Strain	Species	Sero type	Other numbers / details	Country	Date	Source
T08031	*morrisoni*	8a, 8b		Brazil	1991	6
T09010	*tolworthi*	9		U.S.A.	1979	6
T09011	*tolworthi*	9		Iraq	1987	6
T09024	*tolworthi*	9		Indonesia	1991	6
T09034	*tolworthi*	9		Brazil	1992	6
T10001	*darmstadiensi*	10a, 10b		Germany	1967	6
T10003	*darmstadiensi*	10a, 10b		Germany	1975	6
T10016	*darmstadiensi*	10a, 10b		U.S.A.	1982	6
T10017	*darmstadiensi*	10a, 10b		U.S.A.	1982	6
T10018	*darmstadiensi*	10a, 10b		Japan	1982	6
T10024	*darmstadiensi*	10a, 10b		Pakistan	1975	6
T13001	*pakistani*	13		Pakistan	1976	6
T13004	*pakistani*	13		Pakistan	1980	6
T13028	*pakistani*	13		Chile	1993	6
T15001	*dakota*	15		U.S.A.	1983	6
T15006	*dakota*	15		South	1993	6
T18001	*kumamotoensis*	18a, 18b		Japan	1980	6
T18002	*kumamotoensis*	18a, 18b		U.S.A.	1980	6
T18004	*kumamotoensis*	18a, 18b		Iraq	1984	6
4Btt	*tenebrionis*	8a, 8b		U.S.A.		6
CCCT 2259	*israelensis*	14		Brazil		
Ames	*anthracis*		Challenge strain	USA		7
Ames[c]	*anthracis*		Plasmid cured Ames	USA		7
K0610/ A0034	*anthracis*		LSU_34 group A3b Phage resistant ?			7
K4834 / A0039	*anthracis*		LSU_39 group A3a Cattle outbreak.	W. Australia	1994	7
K1340 / A0062	*anthracis*		LSU_62 group A1a	Poland	1962	7
K2478 / A00102	*anthracis*		LSU_102 group B2	Mozambique	1944	7
K2802 / A0248	*anthracis*		LSU_248 group A3d USAMRID-OHIO			7
K3700 / A0267	*anthracis*		LSU_267 group A1b V770-NPI-R US vaccine	USA	1937	7
K1694 / A0462	*anthracis*		LSU_462 group A3b Ames-USAMRID	USA	1932	7
K5135 / A0463	*anthracis*		LSU_463 group A2	Pakistan	1978	7
K2762 / A0465	*anthracis*		LSU_465 group B1	S.W. France	1997	7
K4596 / A0488	*anthracis*		LSU_488 group A4 Vollum CAMR QC strain	U.K.	1997	7
168287_ M	*cereus*		Wound haematoma	UK Glasgow	2003	8
1725601_ V	*cereus*		Calf wound	UK Glasgow	2003	8

Strain	Species	Sero type	Other numbers / details	Country	Date	Source
191560_K	*cereus*		Calf tissue	UK Glasgow	2003	8
R_2955 /03	*cereus*		Blood	UK Bristol	2003	9
R_3039 /03	*cereus*		Blood	UK Leicester	2003	9
R_3098 /03	*cereus*		Blood Post-blood transfusion	UK Hull	2003	9
R_3149 /03	*cereus*		Blood Myeloma patient	UK Manch-ester	2003	9
R_3238 /03	*cereus*		Blood	UK Oxford	2003	9
T01003	*thuringiensis*	1		France	1959	10
T01016	*thuringiensis*	1		U.K.	1962	10
T01058	*thuringiensis*	1		Switzer-land	1964	10
T01100	*thuringiensis*	1		U.S.S.R.	1968	10
T01111	*thuringiensis*	1		Finland	1968	10
T01185	*thuringiensis*	1		Egypt	1976	10
T01194	*thuringiensis*	1		Japan	1976	10
T01213	*thuringiensis*	1		Czecho-slovakia	1979	10
T01217	*thuringiensis*	1		China	1979	10
T01272	*thuringiensis*	1		Belgium	1985	10
T01294	*thuringiensis*	1		Italy	1990	10
T01310	*thuringiensis*	1		Australia	1991	10
T01338	*thuringiensis*	1		Spain	1995	10
T01350	*thuringiensis*	1		Mexico	1999	10
T01354	*thuringiensis*	1		India	2000	10
T04001	*sotto*	4a, 4b		Canada	1970	10
T04003	*sotto*	4a, 4b		France	1968	10
T04012	*sotto*	4a, 4b		U.S.A.	1977	10
T04090	*sotto*	4a, 4b		Pakistan	1982	10
T04205	*sotto*	4a, 4b		U.S.A.	1983	10
T04206	*sotto*	4a, 4b		Czecho-slovakia	1984	10
T04208	*sotto*	4a, 4b		Japan	1985	10
T04214	*sotto*	4a, 4b		Mexico	1988	10
T04232	*sotto*	4a, 4b		Korea	1988	10
T04240	*sotto*	4a, 4b		Indonesia	1991	10
T04245	*sotto*	4a, 4b		Kenya	1992	10
T04251	*sotto*	4a, 4b		China	1994	10
T04254	*sotto*	4a, 4b		Spain	1995	10
T04278	*sotto*	4a, 4b		Pakistan	1997	10
T04280	*sotto*	4a, 4b		Thailand	1998	10
T05a003	*canadensis*	5a, 5c		Chad	1965	10
T05a007	*canadensis*	5a, 5c		Madag-ascar	1965	10
T05a010	*canadensii*	5a, 5c		U.S.A.	1968	10
T05a011	*canadensis*	5a, 5c		Portugal	1972	10
T05a012	*canadensis*	5a, 5c		Czecho-slovakia	1974	10
T05a017	*canadensis*	5a, 5c		Nigeria	1980	10
T05a021	*cnaadensis*	5a, 5c		France	1981	10
T05a024	*canadensis*	5a, 5c		Czecho-slovakia	1982	10

Strain	Species	Sero type	Other numbers / details	Country	Date	Source
T05a027	*canadensis*	5a, 5c		Came-roon	1983	10
T05a031	*canadensis*	5a, 5c		Mexico	1988	10
T05a033	*canadensis*	5a, 5c		Kenya	1989	10
T05a036	*canadensis*	5a, 5c		Korea	1990	10
T05a039	*canadensis*	5a, 5c		Thailand	1995	10
T05a050	*canadensis*	5a, 5c		Benin	1996	10
T05a051	*canadensis*	5a, 5c		Colombia	2000	10
T08003	*morrisoni*	8a, 8b		Canada	1964	10
T08004	*morrisoni*	8a, 8b		U.S.A.	1968	10
T08007	*morrisoni*	8a, 8b		Czecho-slovakia	1976	10
T08011	*morrisoni*	8a, 8b		France	1979	10
T08012	*morrisoni*	8a, 8b		Pakistan	1980	10
T08016	*morrisoni*	8a, 8b		Philip-pines	1985	10
T08018	*morrisoni*	8a, 8b		Germany	1982	10
T08019	*morrisoni*	8a, 8b		Mexico	1985	10
T08021	*morrisoni*	8a, 8b		U.S.A.	1985	10
T08032	*morrisoni*	8a, 8b		Australia	1991	10
T08038	*morrisoni*	8a, 8b		Ghana	1992	10
T08051	*morrisoni*	8a, 8b		Italy	1992	10
T08056	*morrisoni*	8a, 8b		Canada	1992	10
T08109	*morrisoni*	8a, 8b		Nigeria	1995	10
T08121	*morrisoni*	8a, 8b		Peru	1999	10
T09001	*tolworthi*	9		U.K.	1965	10
T09004	*tolworthi*	9		U.S.A.	1967	10
T09006	*tolworthi*	9		U.K.	1975	10
T09009	*tolworthi*	9		U.K.	1976	10
T09016	*tolworthi*	9		Mexico	1988	10
T09020	*tolworthi*	9		Italy	1990	10
T09022	*tolworthi*	9		Australia	1990	10
T09035	*tolworthi*	9		Chile	1993	10
T09043	*tolworthi*	9		Thailand	1994	10
T09045	*tolworthi*	9		Spain	1995	10
T09051	*tolworthi*	9		Australia	1990	10
T09052	*tolworthi*	9		Mexico	1999	10
T09053	*tolworthi*	9		India	2000	10
T10007	*darmstadiensis*	10a, 10b		U.S.S.R.	1981	10
T10008	*darmstadiensis*	10a, 10b		Philip-pines	1980	10
T10011	*darmstadiensis*	10a, 10b		Japan	1981	10
T10014	*darmstadiensis*	10a, 10b		Finland	1981	10
T10028	*darmstadiensis*	10a, 10b		Czecho-slovakia	1982	10
T10029	*darmstadiensis*	10a, 10b		Pakistan	1983	10
T10031	*darmstadiensis*	10a, 10b		Korea	1988	10
T10035	*darmstadiensis*	10a, 10b		Algeria	1993	10
T10036	*darmstadiensis*	10a, 10b		Spain	1995	10

Strain	Species	Sero type	Other numbers / details	Country	Date	Source
T10059	*darmstadiensis*	10a, 10b		China	1995	10
T10063	*darmstadiensis*	10a, 10b		Argentina	1997	10
T14129	*israelensis*	14		Iraq	1987	10
T14240	*israelensis*	14		Colombia	1990	10
T14531	*israelensis*	14		Vietnam	1998	10
4Q1 (BGSC)	*israelensis*	14			1977	11
4Q4 (BGSC)	*israelensis*	14				11
4Q7 (BGSC)	*israelensis* (plasmid free)	14			1987	11
4AA1 (BGSC)	*tenebrionis*	8a, 8b			1989	11
4AB1 (BGSC)	*san diego*	8a, 8b			1987	11

APPENDIX B

Alleles for MLST Alleles shown are from *B. anthracis* (Ames)

>glp-1 381 bases
GTAGCGGTTGCAGCGTACGTGGTTGGATCAATTAGTGGGGCACATTTGAATCCAGCTTTG
ACGATAGGATTAGCATTTAAGGGAGCGTTCCCATGGGGTGACGTACCTGGTTATATCGCA
GCACAAATGATTGGGGCAATTATCGGGGCAGTTATCGTATATTTACATTACTTACCACAC
TGGAAAGAAACAGAAGATCCAGGAACAAAGTTAGGTGTATTTGCAACAGGTCCAGCAATT
CCGAACACATTTGCAAACCTTTTAAGTGAAATGATTGGGACATTCGTTTTAGTATTTGGT
ATATTAGCAATTGGTGCAAATAAATTTGCAGATGGATTAAATCCATTTATCGTAGGTTTC
TTAATTGTAAGTATTGGTTTA

>gmk-1 504 bases
GTTCTTTCAGGACCTTCTGGTGTTGGGAAAGGAACGGTTCGAAAAGAGCTGTTTAGTCAT
GAGGATACACGTTTTCAGTACTCCATTTCAGTAACGACACGTAAGCCACGTGAAGGTGAA
GTGGATGGTGTGGATTATTTCTTTAAAGAGAGAGAAGAATTTGAGGAAATGATTCGTAAT
GAAAAATTACTTGAGTGGGCTGAGTTCGTAGGTAATTATTACGGAACACCAATTGACTAT
GTTGAAAAAACATTGCAAGAAGGAAAAGATGTATTCTTAGAAATTGAAGTGCAAGGAGCA
ATTCAAGTTAAGAAAGCTTTCCCAGAAGGTGTATTTATTTTCTTAGCACCTCCAAGTTTA
TCTGAACTAAAGAACCGAATTGTCGGTCGTGGTACAGAAACTGAAGATGTTATTGAAAAT
CGTTTAACGGTAGCGAAAGAAGAAATTGATATGATGGACGCTTACGACTACGTTGTAGAA
AACGACCAAGTTGAACTAGCTTGT

>ilv-1 393 bases
GAAGAAGGTGGATTACGTATATTAAAAGGAAACCTTGCGAAAGACGGAGCAGTTATTAAA
AGCGGGGCAACTGAAGTAAAACGATTTGAAGGACCTTGTGTTATTTTTAATTCACAAGAT
GAGGCGCTTGCCGGCATTATGCTTGGGAAGGTTAAGAAAGGAGATGTAGTTGTTATTCGT
TATGAAGGACCAAGAGGCGGTCCTGGTATGCCGGAAATGTTAGCACCAACGTCAGCGATT
GCTGGCATGGGATTAGGTGCAGATGTTGCGTTATTAACCGATGGTCGTTTCTCTGGTGCT
TCACGTGGTATTTCAGTAGGTCATATTTCGCCAGAAGCAGCTGCGGGCGGAACGATTGCA
CTTCTTGAACAAGGGGATATCGTTTGTATCGAT

>pta-1 414 bases
AATTTAACATTAGCAGGCGTTGATATTTACGATCCAGCTACATACGAAGAAATGGATGCA
ATGGTAGCATCTTTCGTTGAACGCCGTAAAGGTAAAGCAACAGAAGAAGACGCTCGCAAA
ATCCTTAAAGACGAAAATTACTTCGGTACAATGCTTGTATACATGGGCAAAGCACACGGT
TTAGTAAGTGGTGCAGCTCACTCTACAGCTGATACAGTTCGTCCAGCACTTCAAATTATT
AAAACAAAACCAGGCGTTACAAAAACTTCTGGCGTATTCATCATGGTACGTGAAGAAGAG
AAATATGTATTTGCTGATTGCGCAATTAACATTGCACCAAATAGCCAAGATTTAGCTGAA
ATTGGTATCGAGAGTGCGAAAACTGCTGAACTATTCGGTATTGACCCACGCGTT

>pur-1 348 bases
AAACGTAAACTAGCAGCGAAAGTATTCCGTCATACAGCAGCGTATGATGCGTTAATTTCT
AACTACTTAACAGAGCAAATGGGTGAAGAAAGTCCAGAAACATTAACTGTGACATTTGAG
AAAAAGCAAGACTTACGTTATGGCGAGAACCCACATCAAAAAGCAACTTTCTATAAAGCG
CCATTCGCAGCAACTTCTTCTGTTGCATACGCAGAACAATTACATGGCAAGGAATTATCG
TATAACAATATTAACGATGCAGATGCAGCGCTTAGTATCGTGAAAGAATTTACAGAACCA
GCAGTAGTAGCGGTAAAACATATGAATCCATGTGGTGTTGGAGTAGGT

>pyc-1 363 bases
TACTACGCACCGTTTGAAAGTGGTATGAACGCTCCTCATACAGAGGTATATATGCACGAA
ATGCCAGGCGGACAGTACAGTAATTTACAACAACAAGCGAAAGTGGTTGGCTTAGGCGAT
CGCTTCGATGAAGTGAAAGTAATGTATCGCCGTGTGAATGATATGTTTGGTGATATTGTT
AAAGTAACGCCATCATCTAAAGTTGTCGGTGATATGGCATTATTCATGGTTCAAAATCAC
TTAACAGAACAAGATGTTTTAGAGCGTGGACATTCTATGGACTTCCCAGGATCTGTTGTT
GAAATGTTCTCTGGAGACTTAGGACAACCATATGGTGGTTTCCCGAAAAAGTTACAAGAA
ATT

>tpi-1 435 bases
GAGTCAGTAAACAAAAAGACTATTGCAGCATTTGAACATGGTTTAACACCAATCGTATGT
TGTGGTGAGACTTTAGAAGAGCGCGAAAGCGGAAAAACATTTGATCTAGTAGCAGGTCAA
GTGACAAAAGCACTTGCAGGTTTAACAGAAGAGCAAGTTAAAGCAACTGTTATCGCTTAT
GAGCCAATCTGGGCTATCGGTACAGGTAAATCTTCTTCTTCTGCAGATGCAAACGAAGTA
TGTGCGCACATCCGTAAAGTTGTTGCAGAAGTTGTTTCTCCAGCTGCTGCAGAAGCTGTT
CGTATCCAATACGGCGGTAGCGTAAAACCAGAAAACATTAAAGAGTACATGGCACAATCT
GACATCGACGGCGCTTTAGTTGGCGGTGCTAGCTTAGAGCCTGCTTCGTTCTTAGGTCTT
CTGGGGGCGGTAAAA

APPENDIX C

Genbank submissions

A set of 56 allele sequences, comprising seven alleles from eight strains, were submitted to GenBank as a 'POPSET', with the following accession numbers: -

Table 5.2 GenBank accession numbers

allele	*Accession number*
glpF	AY729746 – AY729753
gmk	AY729754 – AY729761
ilvD	AY729762 – AY729769
pta	AY729770 – AY727777
pur	AY729778 – AY729785
pycA	AY729786 – AY729793
tpi	AY729794 – AY729801

The alleles submitted to GenBank were from the following strains:-

B. anthracis group A (Ames)

B. anthracis group B

B. cereus F4810/72 emetic strain

B. thuringiensis sotto T04002

B. thuringiensis kurstaki T03a001

B. mycoides WSBC 10277

B. weihenstephanensis WSBC 10364

B. pseudomycoides WS 3118

APPENDIX D

Table showing all strains and the analytical methods applied to each one.

MLST – Multi-Locus Sequence Typing

RAPD – Random Amplified Polymorphic DNA.

Crystal protein genes – Analysis of Crystal toxin genes by PCR**.**

Enterotoxin Genes – Analysis of enterotoxin genes (Hbl and Nhe) by PCR.

Starch Hydrolysis – Test for ability to hydrolyse starch.

SDS-Page – Analysis of Crystal proteins by SDS-Page.

Table 5.3 Analytical Methods applied to strains

	MLST	RAPD	Crystal protein genes	Entero toxin genes	Starch hydrolysis	SDS-Page
B. cereus						
S 57	X	X	X	X	X	X
S 58	X	X	X	X	X	X
S 59	X	X	X	X	X	X
S 61				X	X	
S 62				X	X	
S 63				X	X	
S 64	X	X		X	X	
S 66	X	X		X	X	
S 67				X	X	
S 68				X	X	
S 70	X	X	X	X	X	
S 71	X	X		X	X	
S 72				X	X	
S 73		X		X	X	
S 74	X	X		X	X	
S 75				X	X	
S 76				X	X	
S 77				X	X	
S 78	X			X	X	
S 79	X			X	X	
S 86	X	X		X	X	
S 363	X	X		X	X	
S 366	X	X		X	X	
S 710	X	X		X	X	
AH 75	X	X		X	X	
AH 181	X	X		X	X	
AH 182	X	X		X	X	
AH 517				X	X	
AH 563				X	X	
AH 581				X	X	

	MLST	RAPD	Crystal protein genes	Entero toxin genes	Starch hydrolysis	SDS-Page
AH 621	X	X		X	X	
AH 636				X	X	
AH 647	X			X	X	
AH 651				X	X	
AH 656				X	X	
AH 662				X	X	
AH 671				X	X	
AH 681				X	X	
AH 684	X			X	X	
AH 685				X	X	
AH 686				X	X	
AH 691				X	X	
AH 693				X	X	
AH 821	X	X		X	X	
AH 923	X	X		X	X	X
WSBC 10028	X			X	X	
WSBC 10249	X			X	X	
WSBC 10312	X			X	X	
WSBC 10360	X	X		X	X	
WSBC 10277	X	X		X	X	
WS 3118	X			X	X	
WS 2629				X	X	
WS 2734				X	X	
WSBC 28002				X	X	
WSBC 10202	X	X		X	X	
WSBC 10204				X	X	
WSBC 10364	X			X	X	
F4810/72	X	X		X	X	
F3942/87	X	X		X	X	
F3080B/87	X	X		X	X	
M21	X			X	X	
H 5				X	X	
TSP11	X			X	X	
BM-VIIIP1				X	X	
Cal3	X			X	X	
TSP10	X			X	X	
SPS 2	X	X		X	X	
TSP8	X			X	X	
K8	X			X	X	
m1282				X	X	
m1291				X	X	
m1550	X			X	X	
m1558				X	X	
m1552	X			X	X	
m1576	X			X	X	
m1584				X	X	
m1592				X	X	
m90 / 4				X	X	
m218 / 3				X	X	

	MLST	RAPD	Crystal protein genes	Entero toxin genes	Starch hydrolysis	SDS-Page
m1557				X	X	
m213 / 5				X	X	
m166 / 5				X	X	
m1278	X	X		X	X	
m1280	X	X		X	X	
m1292	X	X		X	X	
m1293	X	X		X	X	
m1545	X	X		X	X	
m1564	X	X		X	X	
m1568						
m1574						
m95/5						
m11/4						
m96/1						
m92/5						
m221/2						
B. thuringiensis						
T01001	X	X	X	X	X	
T01015	X	X	X	X	X	
T01022	X	X	X	X	X	
T01246	X	X	X	X	X	
T01326	X	X	X	X	X	
T03a001	X	X		X	X	X
T03a075	X	X		X	X	
T03a172	X	X		X	X	
T03a287	X			X	X	
T03a361	X		X	X	X	X
T04002	X	X	X	X	X	
T04016	X	X		X	X	
T04024	X	X		X	X	
T04236	X	X		X	X	
T04b001	X			X	X	
T04b012	X			X	X	
T04b054	X			X	X	
T04b060	X			X	X	
T04b073	X			X	X	
T05005	X			X	X	
T05033	X			X	X	
T05114	X			X	X	
T05125	X			X	X	
T05a001	X	X		X	X	
T05a015	X	X		X	X	
T05a019	X	X		X	X	
T05a030	X	X		X	X	
T06002			X	X	X	
T06007	X			X	X	
T06010	X			X	X	
T07033	X			X	X	

	MLST	RAPD	Crystal protein genes	Entero toxin genes	Starch hydrolysis	SDS-Page
T07058	X			X	X	
T07146	X			X	X	
T07180	X	X		X	X	
T07196	X			X	X	
T08001	X	X	X	X	X	
T08009	X	X	X	X	X	
T08012	X	X	X	X	X	
T08023	X	X	X	X	X	
T08025	X	X	X	X	X	
T08031	X	X	X	X	X	
T09010	X	X	X	X	X	
T09011	X	X	X	X	X	
T09024	X	X	X	X	X	
T09034	X	X	X	X	X	
T10001	X	X		X	X	
T10003	X	X		X	X	
T10016	X	X		X	X	
T10017	X	X		X	X	
T10018	X	X		X	X	
T10024	X	X	X	X	X	
T13001	X			X	X	
T13004	X			X	X	
T13028	X	X		X	X	
T15001	X			X	X	
T15006	X			X	X	
T18001	X			X	X	
T18002	X			X	X	
T18004	X	X	X	X	X	
T37001		X				
4btt	X	X	X	X	X	
976	X	X	X	X	X	
B. anthracis						
Ames	X	X				
Delta Ames	X					
LSU_34	X					
LSU_39	X					
LSU_62	X					
LSU_102	X					
LSU_248	X					
LSU_267	X					
LSU_462	X					
LSU_463	X					
LSU_465	X					
LSU_488	X					
B. cereus						
168287_M	X	X		X		
1725601_V	X	X		X		
191560_K	X	X		X		
R_2955	X	X		X		
R_3039	X	X		X		

	MLST	RAPD	Crystal protein genes	Entero toxin genes	Starch hydrolysis	SDS-Page
R_3098	X	X		X		
R_3149	X	X		X		
R_3238	X	X		X		
Additional *B. thuringiensis*						
T01003		X	X			
T01016		X	X			
T01058	X	X	X			
T01100		X	X			
T01111		X	X			
T01185		X	X			
T01194		X	X			
T01213		X	X			
T01217		X	X			
T01272		X	X			
T01294		X	X			
T01310		X	X			
T01338		X	X			
T01350		X	X			
T01354		X	X			
T04001		X				
T04003		X				
T04012		X				
T04090		X				
T04205		X				
T04206		X				
T04214		X				
T04232		X				
T04240		X				
T04245		X				
T04251		X				
T04254		X				
T04278		X				
T04280		X				
T05a003		X				
T05a007		X				
T05a010		X				
T05a011		X				
T05a012		X				
T05a017		X				
T05a021		X				
T05a024		X				
T05a027		X				
T05a031		X				
T05a033		X				
T05a036		X				
T05a039		X				
T05a050		X				
T05a051		X				

	MLST	RAPD	Crystal protein genes	Entero toxin genes	Starch hydrolysis	SDS-Page
T08003	X	X	X			
T08004		X	X			
T08007		X	X			
T08011		X	X			
T08012		X	X			
T08016	X	X	X			
T08018		X	X			
T08019	X	X	X			
T08021		X	X			
T08032		X	X			
T08038	X	X	X			
T08051		X	X			
T08056		X	X			
T08109		X	X			
T08121	X	X	X			
T09001		X	X			
T09004		X	X			
T09006		X	X			
T09009		X	X			
T09016	X	X	X			
T09020	X	X	X			
T09022		X	X			
T09035		X	X			
T09043	X	X	X			
T09045		X	X			
T09051		X	X			
T09052	X	X	X			
T09053		X	X			
T10007		X				
T10008		X				
T10011		X				
T10014		X				
T10028		X				
T10029		X				
T10031		X				
T10035		X				
T10036		X				
T10059		X				
T10060		X				
T10063		X				
T14129		X				
T14240		X	X			
T14531	X	X	X			
4Q1 (BGSC)		X				
4Q4 (BGSC)		X				
4Q7 (BGSC)		X				

	MLST	RAPD	Crystal protein genes	Entero toxin genes	Starch hydrolysis	SDS-Page
4AA1 (BGSC)		X	X			
4AB1 (BGSC)		X				

Some of the enterotoxin tests for strains of *B. thuringiensis* were carried out by A, Gaviria-Rivera as part of her Ph. D project and are included here for completeness**.**

APPENDIX E

Allele trees for 88 STs

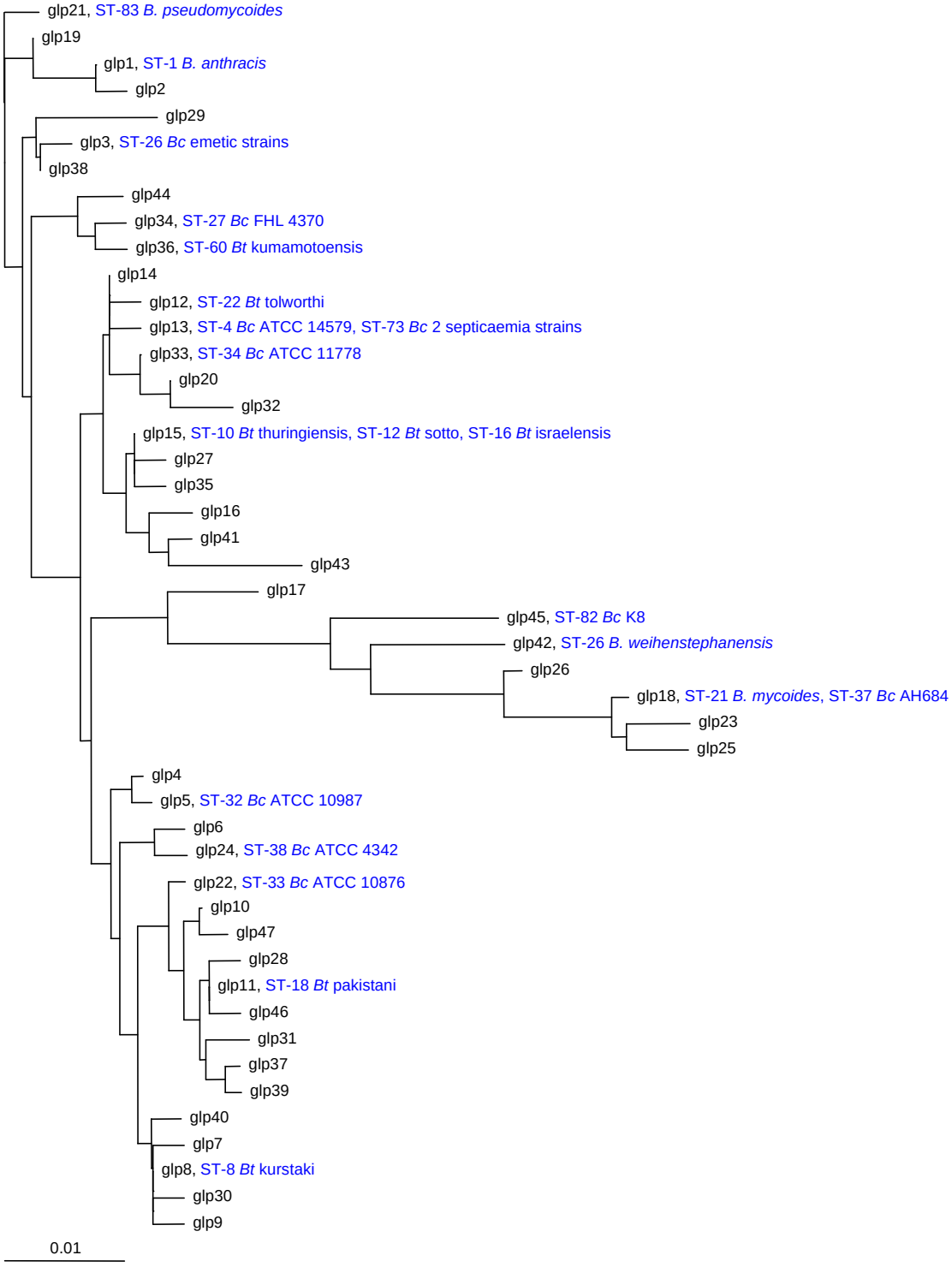


Figure 5.1 Neighbour Joining tree for 47 ***glp*** alleles from 88 Sequence Types

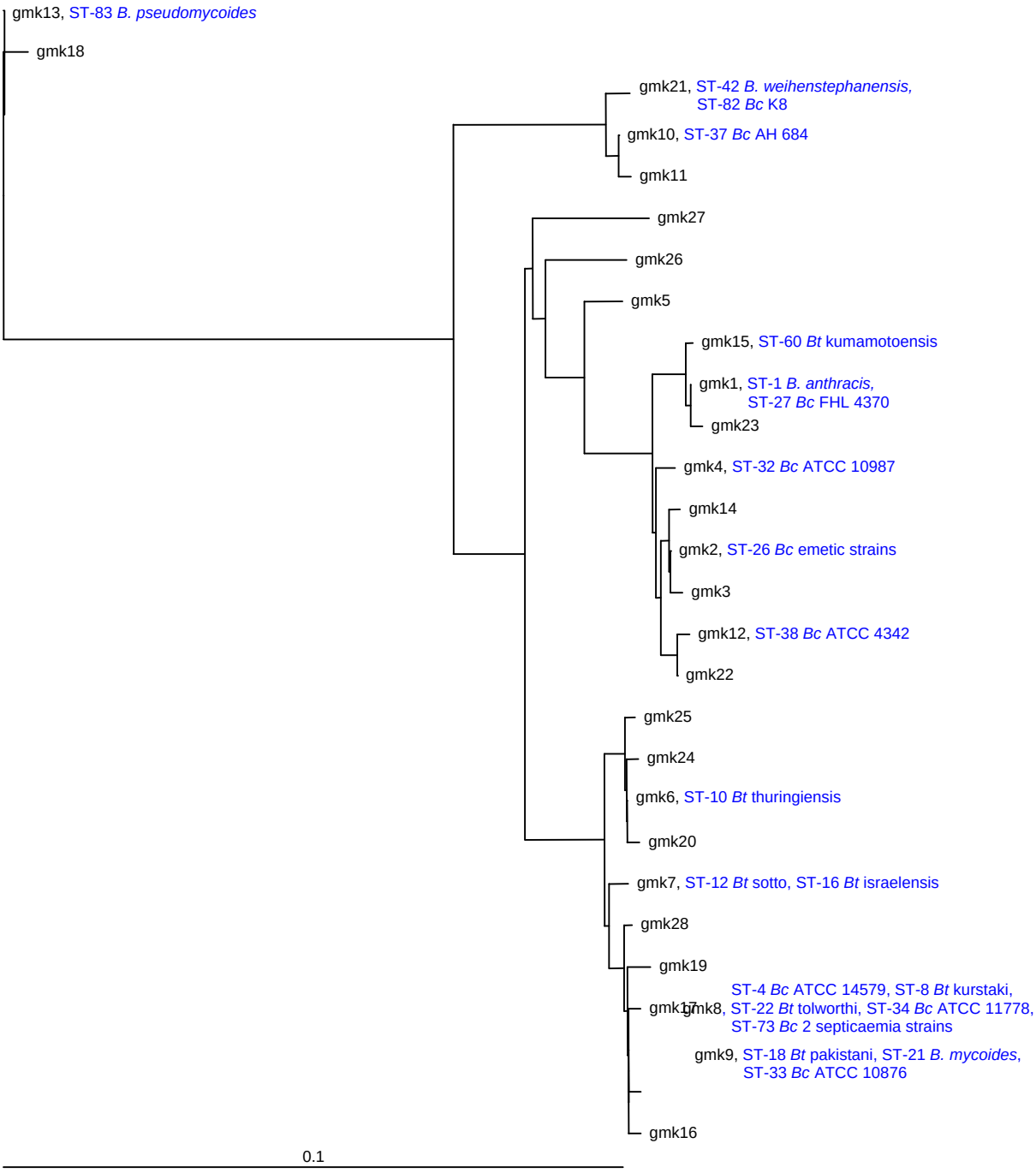


Figure 5.2 Neighbour Joining tree for 28 ***gmk*** alleles from 88 Sequence Types

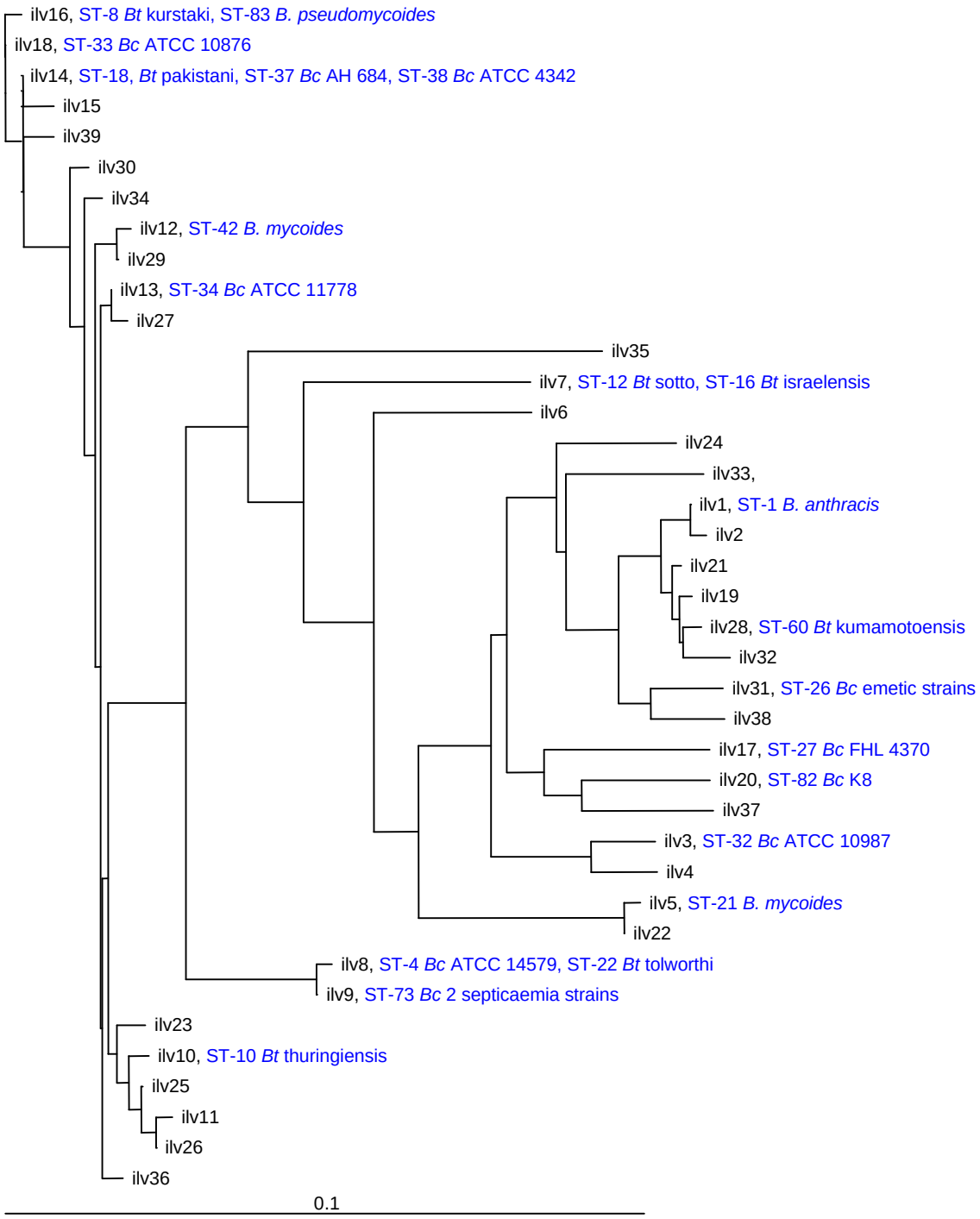


Figure 5.3 Neighbour Joining tree of 39 ***ilv*** alleles from 88 Sequence Types.

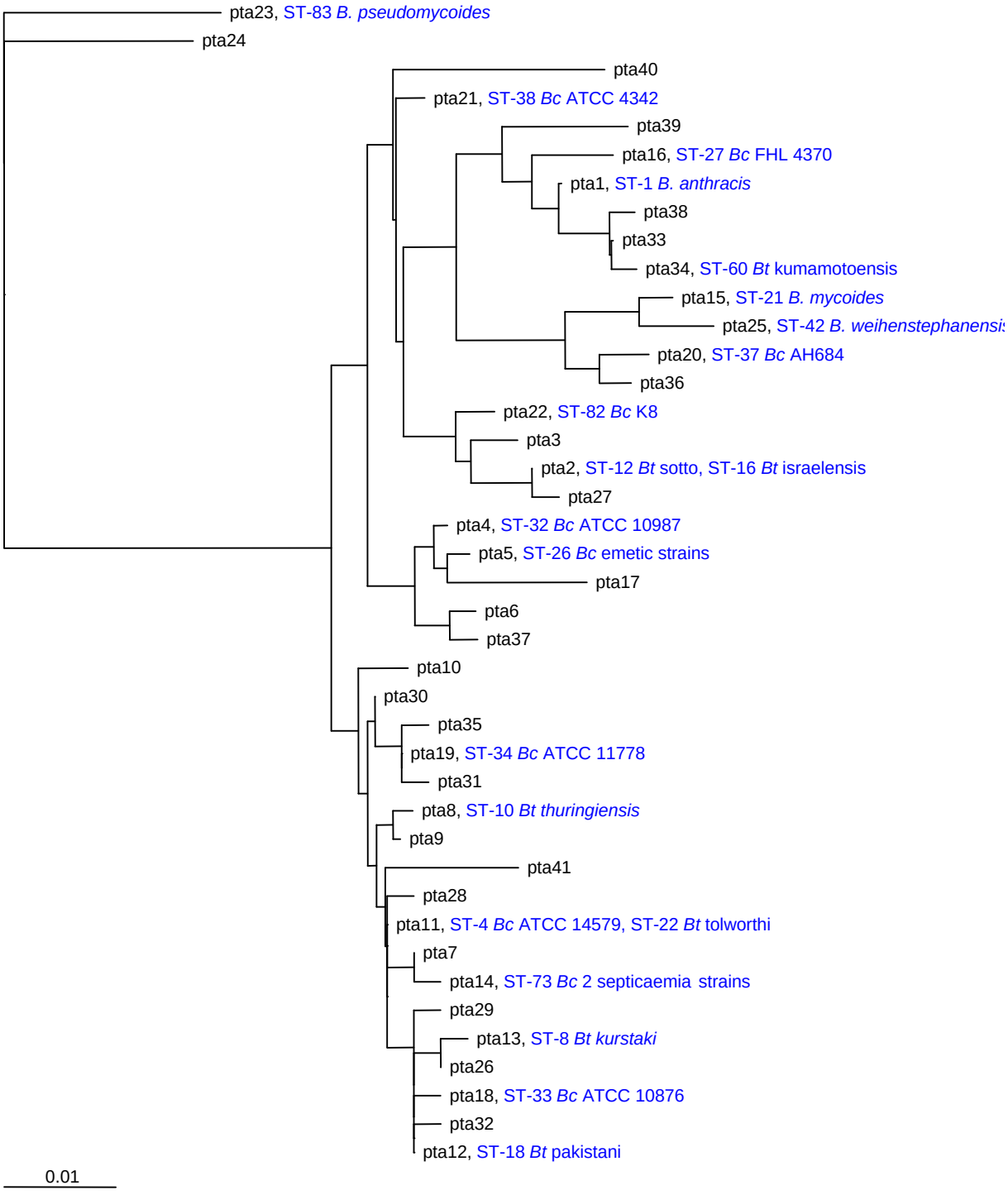


Figure 5.4 Neighbour Joining tree of 41 ***pta*** alleles from 88 Sequence Types.

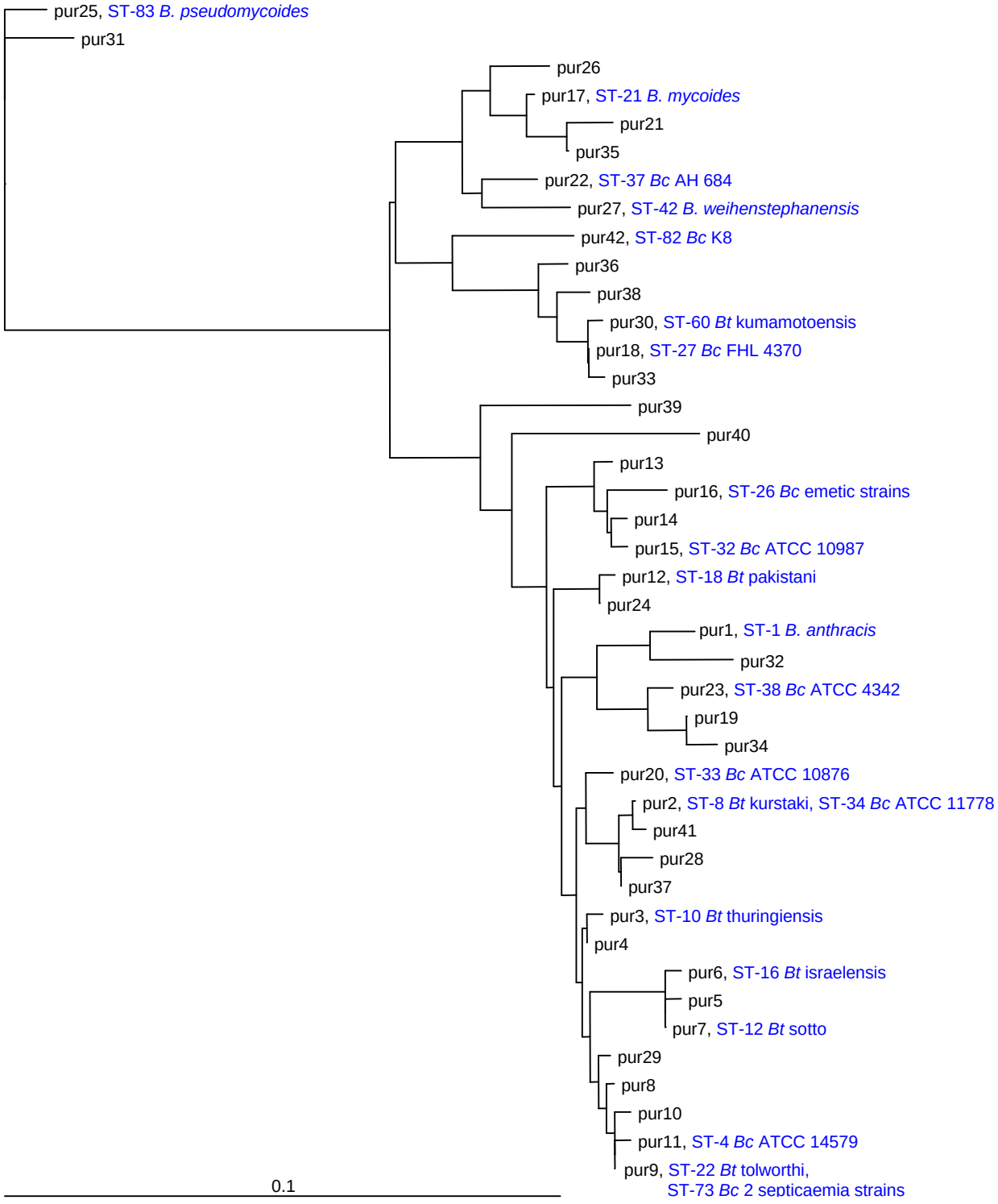


Figure 5.5 Neighbour Joining tree of 42 ***pur*** alleles from 88 Sequence Types

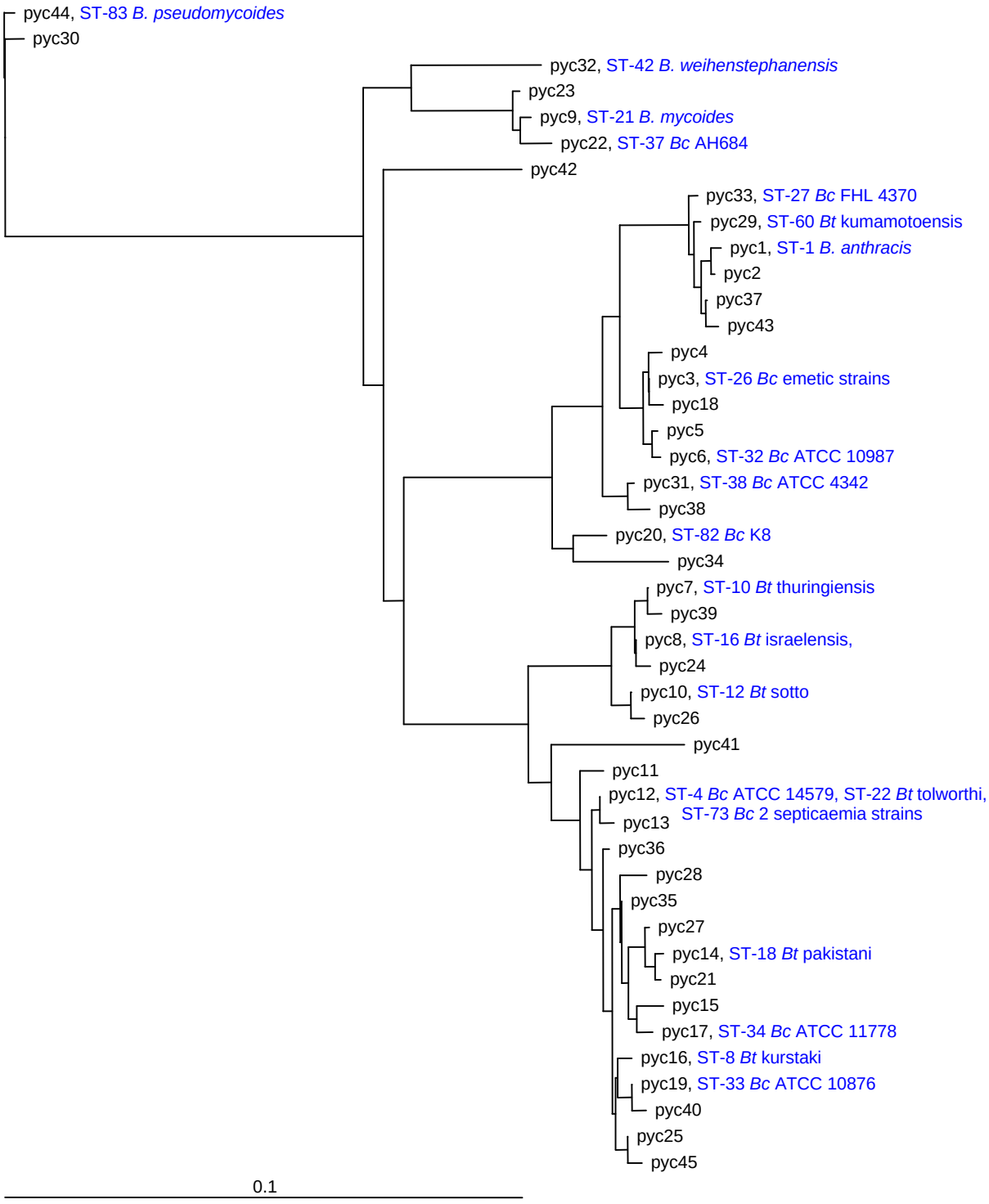


Figure 5.6 Neighbour Joining tree of 45 ***pyc*** alleles from 88 Sequence Types

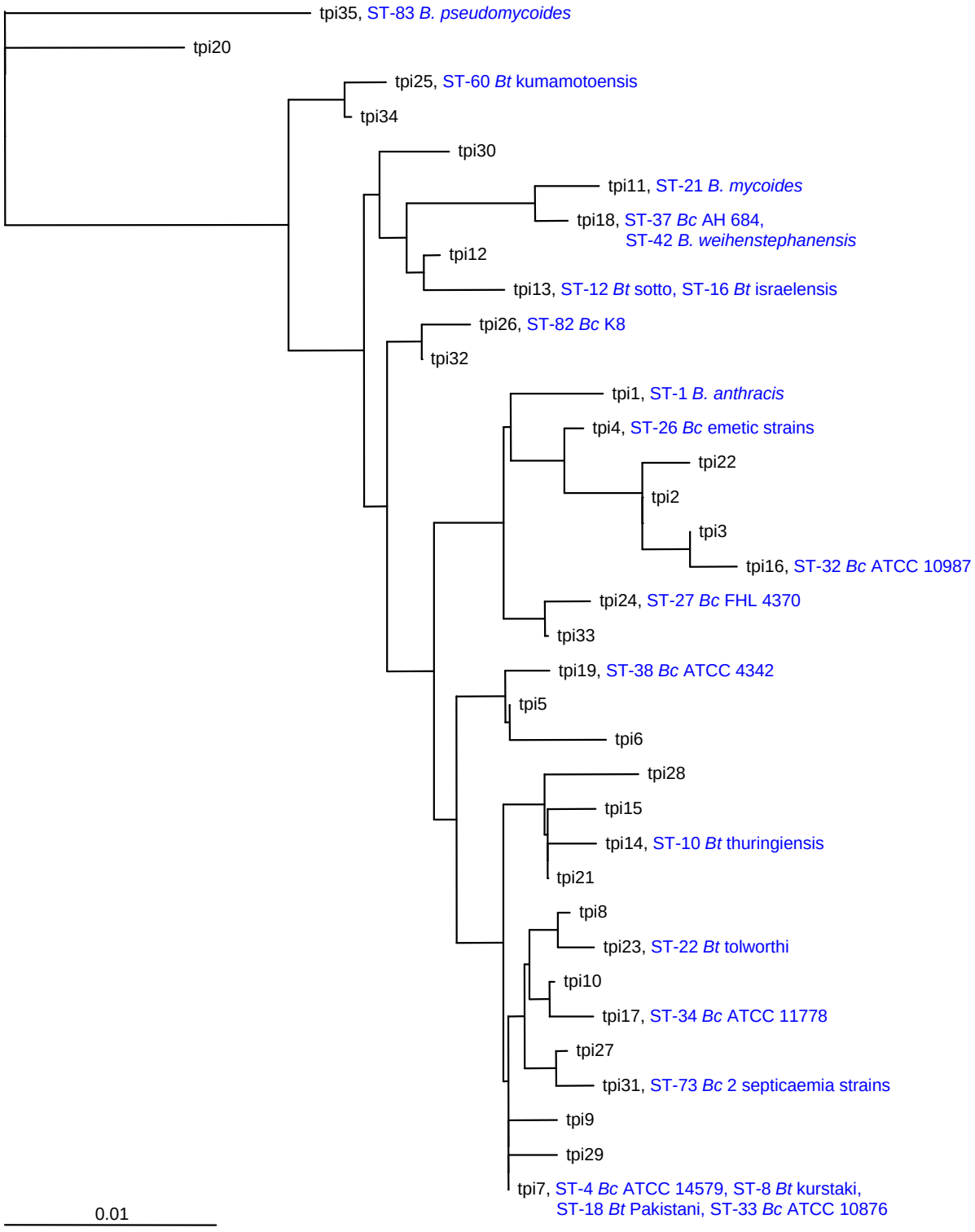


Figure 5.7 Neighbour Joining tree of 35 ***tpi*** alleles from 88 Sequence Types

REFERENCES

Aanensen, D.M., Spratt, B.G. 2005. The multilocus sequence typing network: mlst.net. Nucleic Acids Res**. 33:** W728-W733.

Abrami L, Reig N, van der Goot FG. 2005 Anthrax toxin: the long and winding road that leads to the kill. Trends Microbiol. **13:** 72-8.

Achtman, M., Zurth, K., Morelli, C., Torrea, G., Guiyoule, A., Carniel, E. 1999. *Yersinia pestis,* the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis.* Proc. Natl. Acad. Sci. USA **96:** 14043-14048.

Agaisse, H., Gominet, M., Økstad, O.E., Kolstø, A-B., Lereclus, D. 1999. PlcR is a pleiotropic regulator of extracellular virulence factor gene expression in *Bacillus thuringiensis.* Mol. Microbiol. **32:** 1043-1053.

Agata, N., Ohta, M., Mori, M., Isobe, M., 1995. A novel dodecadepsipeptide, cereulide, is an emetic toxin of Bacillus cereus. FEMS Mirobiol. Let. **129:** 17-20.

Agata, N., Ohta, M., Mori, M. 1996. Production of an emetic toxin, cereulide, is associated with a specific class of *Bacillus cereus.* Curr. Microbiol. **33:** 67-69.

Agata, N., Ohta, M., Yokoyama, K. 2002. Production of *Bacillus cereus* emetic toxin (cereulide) in various foods. Int. Jour. Food Microbiol.**73:** 23-27.

Andersson, M.A., Mikkola, R., Helin, J., Andersson, M.C., Salkinoja-Salonen, M. 1998. A novel sensitive bioassay for detection of *Bacillus cereus* emetic toxin and related depepsipeptide ionophores. Appl. Environ. Microbiol. **64:** 1338-1343.

Aronson, A.I., Shai, Y. 2001. Why *Bacillus thuringiensis* insecticidal toxins are so effective: unique features of their mode of action. FEMS Microbiol. Lett. **195:** 1-8.

Ash, C., Farrow, J.A.E., Dorsch, M., Stackebrandt, E., Collins, M.D. 1991. Comparative analysis of *Bacillus anthracis, Bacillus cereus,* and related species on the

basis of reverse transcriptase sequencing of 16S rRNA. Int. J. System. Bacteriol. **41:** 343-346.

Ash, C., Collins, M.D. 1992. Comparative analysis of 23S ribosomal RNA gene sequences of *Bacillus anthracis* and emetic *Bacillus cereus* determined by PCR-direct sequencing. FEMS Microbiol. Lett. **94:** 75-80.

Baker, L., Brown, T., Maiden, M.C.J., Drobniewski, F. 2004. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. Emerg. Infect. Dis. **10:** 1568-1577.

Barker, M., Thakker, B., Priest, F.G. 2005**.** Multilocus sequence typing reveals that *Bacillus cereus* strains isolated from clinical infections have distinct phylogenetic origins. FEMS Microbiol. Lett. **245:** 179-184.

Battisti, L., Green, B.D., Thorne, C.B. 1985. Mating system for transfer of plasmids among *Bacillus anthracis, Bacillus cereus* and *Bacillus thuringiensis*. J. Bacteriol. **162:** 543-550.

Bavykin, S.G., Lysov, Y.P., Zkhariev, V., Kelly, J.J., Jackman, J., Stahl, D.A., Cherni, A. 2004. Use of 16S rRNA, 23S rRNA, and *gyrB* gene sequence analysis to determine phylogenetic relationships of *Bacillus cereus* group microorganisms. J. Clin. Microbiol. **42:** 3711-3730.

Beecher, D.J., Macmillan, J.D. 1990. A novel bicomponent hemolysin from *Bacillus cereus*. Infect. Immun. **58:** 2220-2227.

Beecher, D.J., Macmillan, J.D. 1991. Characterization of the components of hemolysin BL from *Bacillus cereus.* Infect. Immun. **59:** 1778-1784.

Beecher, D.J., Wong, A.C.L. 2000. Cooperative, synergistic and antagonistic haemolytic interactions between haemolysin BL, phosphatidylcholine phospholipase C and sphingomyelinase from *Bacillus cereus*. Microbiol. **146:** 3033-3039.

Blackwood, K. S., Turenne, C. Y., Harmsen, D., Kabani, A. M. 2004. Reassessment of sequence-based targets for identification of *Bacillus* species. J Clin Microbiol. 2004 **42:** 1626-30.

Bougnoux, M.-E., Morand, S., d'Enfert, C. 2002. Usefulness of Multilocus Sequence Typing for Characterization of Clinical Isolates of *Candida albicans.* J. Clin. Microbiol. **40:** 1290-1297.

Boyd, E.F., Nelson, K., Wang, F-S., Whittam, T.S., Selander, R.K. 1994. Molecular genetic basis of allelic polymorphism in malate dehydrogenase (*mdh)* in natural populations of *Escherichia coli* and *Salmonella enterica.* Proc. Natl. Acad. Sci. USA **91:** 1280-1284.

Bravo, A. 1997. Phylogenetic Relationships of *Bacillus thuringiensis* d-Endotoxin Family Proteins and Their Functional Domains. J. Bacteriol. **179:** 2793-2801.

Brenner, D. J., Staley, J.T., Kreig, N.R. 2001. Classification of prokaryotic organisms and the concept of bacterial speciation, p. 27-31. In G.M. Garrity (ed), Bergey's Manual of Systematic Bacteriology, 2nd ed. Springer, New York, N.Y.

Brousseau, R., Saint-Onge, A., Prefontaine, G., Masson, L., Cabana, J. 1993. Arbitrary primer polymerase chain reaction, a powerful method to identify *Bacillus thuringiensis* serovars and strains. Appl. Environ. Microbiol. **59**:114-119.

Brown, T.A. 2002. Genomes. 2nd ed. Biosis Scientific Publishers Ltd. 420-430

Carlson, C.R., Johansen, T., Kolsta, A.B. 1996. The chromosome map of *Bacillus thuringiensis* subsp. *canadensis* HD224 is highly similar to that of the *Bacillus cereus* type strain ATCC 14579. FEMS Microbiol. Let. **141:** 163-167.

Cherif. A., Borin. S., Rizzi. A., Ouzari. H., Boudabous. A., Daffonchio. D. 2003. *Bacillus anthracis* **d**iverges from related clades of the *Bacillus cereus* group in 16S-23S ribosomal DNA intergenic transcribed spacers containing tRNA Genes. Appl. Environ. Microbiol. **69:** 33-40.

Cohan, F. M. 2002. What are bacterial Species ? Annu. Rev. Microbiol. **56:** 457-487.

Crickmore, N., Zeigler, D.R., Feitelson, J., Schnepf, E., Van Rie, J., Lereclus, D., Baum, J., Dean, D.H. 1998. Revision of the nomenclature for the *Bacillus thuringiensis* pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. **62:** 807-813.

Cooper J. E., Feil, E. J. 2004. Multilocus sequence typing - what is resolved ? Trends Microbiol. **12:** 373-7.

Daffonchio, D., Borin, S., Frova, G., Gallo, R., Mori, E., Fani, R., Sorlini, C. 1999. A Randomly amplified polymorphic DNA marker specific for the *Bacillus cereus* group is diagnostic for *Bacillus anthracis*. Appl. Environ. Microbiol. **65:** 1298-1303.

Damgaard, P. H. 1995. Diarrhoeal enterotoxin production by strains of *Bacillus thuringiensis* isolated from commercial *Bacillus thuringiensis*-based insecticides. FEMS Immunol Med Microbiol. **2:** 245 – 250.

Damgaard, P.H., Granum P.E., Bresciani J., Torregrossa M.V., Eilenberg J., and Vealentino L. 1997. Characterization of *Bacillus thuringiensis* isolated from infections in burn wounds. FEMS Immunol. Med. Microbiol. **18:** 47-53.

De Barjac, H ., Bonnefoi, A. 1962. Essai de classification biochimique et serologique de
24 souches de *Bacillus* du type *B. thuringiensis*. Entomophaga. **7:** 5-31.

Debro, L. Fitz-James, P. C., Aronson. A. 1986. Two different parasporal inclusions are produced by *Bacillus thuringiensis* subsp. *finitimus***.** J. Bacteriol. **165:** 258–268.

Dingle, K.E., Colles, F.M., Wareing, R.A., Ure, R., Fox, A.J., Bolton, F.E., Bootsma, H.J., Willems, R.J.L., Urwin, R., Maiden, M.C.J. 2001. Multilocus sequence typing system for *Campylobacter jejuni.* J. Clin. Microbiol. **39:** 14-23.

Dingle, K.E., Colles, F.M., Ure, R., Jaap, A., Wagenaar, J.A., Dulm, B., Bolton, F.E., Fox, A.J., David, R.A., Wareing, R.A., Maiden, M.C.J. 2002. Molecular

characterization of *Campylobacter jejuni* clones: A basis for epidemiologic investigation. Emerg. Infect. Diseases. **8:** 949-955.

Drobniewski, F.A. 1993. *Bacillus cereus* and related species. Clin. Microbiol. Rev. **6:**324-338.

Dubnau, D. 1991. Genetic competence in *Bacillus subtilis*. Microbiol. Rev. **55:** 395–424.

Dykhuizen, D.E., Green, L. 1991. Recombination in *Escherichia coli* and the definition of biological species. J. Bacteriol. **173:** 7257-7268.

Dykhuizen, D.E. 2000. *Yersinia pestis*: an instant species? Trends Microbiol**. 8:** 296-298.

Ehling-Schulz, M., Svensson, B., Guinebretiere, M-H., Lindbäck, T., Anderssom, M., Schulz, A., Fricker, M., Christiansson, A., Granum, P.E., Märtlbauer, E., Nguyen-The, C., Salkinoja-Salonen, M., Scherer, S. 2005. Emetic toxin formation of *Bacillus cereus* is restricted to a single evolutionary lineage of closely related strains. Microbiol. **151:** 183-197.

Embley, T.M. 1991. The linear PCR reaction: a simple and robust method for sequencing rRNA genes. Lett. Appl. Microbiol. **13:** 171-174.

Enright, M.C., Spratt B.G. 1998. A multilocus sequence typing scheme for *Streptococcus pmeumoniae* : identification of clones associated with serious invasive disease. Microbiol. **144:** 3049-3060.

Enright, M.C., Day, N.P.J., Davies, C.E., Peacock, S.J.,Spratt, B.G. 2000. 'Multilocus Sequence Typing for Characterization of Methicillin-Resistant and Methicillin Susceptible Clones of *Staphylococcus aureus'*. J. Clin. Microbiol. **38:** 1008-1015.

Enright, M.C., Robinson, D.A., Randle, G., Feil, E.J., Grundmabnn, H., Spratt,B. 2002. The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). Proc. Natl. Acad. Sci. USA **99:** 7687-7692.

European Bioinformatics Institute. Protein database 'PDBsum' and Secondary Annotated Structure database 'SAS'.

Falush. D., Kraft, C., Taylor, N.S., Correa, P., Fox, J.G., Achtman, M. Suerbaum, S. 2001. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size and minimal age. Proc. Natl. Acad. Sci. USA **98:** 15056-15061.

Feil, E.J., Maiden, M.C.J., Achtman, M., Spratt, B.G. 1999. The Relative Contributions of Recombination and Mutation to the Divergence of Clones of *Neisseria meningitides.* Mol. Biol. Evol. **16:** 1496-1502.

Feil, E.J., Enright, M. C., Spratt, B.G. 2000. Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between *Neisseria meningitides* and *Streptococcus pneumoniae*. Res. Microbiol. **151:** 465 – 469.

Feil, E.J., Holmes, E.C., Bessen,D.E., Chan, M-S., Day, N.P.J., Enright, M.C., Goldstein, R., Hood, D.W., Kalia, A., Moore, C.E., Zhou, J., Feil. E.J., Spratt, B.G. 2001. Recombination within natural populations of pathogenic bacteria: Short-term empirical estimates and long-term phylogenetic consequences. Proc. Natl. Acad. Sci. USA **98:** 182-187

Feil, E.J., Spratt, B.G. 2001. Recombination and the population structures of bacterial pathogens. Annu. Rev. Microbiol. **55:** 561-590.

Feil, E.J., Cooper, J.E., Grundmann, H., Ashley Robinson, D., Enright, M.C., Berendt, T., Peacock, S.J., Maynard Smith, J., Murphy, M., Spratt, B.G., Moore, C.E., Day, N.P.J. 2003. How Clonal is *Staphylococcus aureus* ? J. Bacteriol. **185**: 3307-3316

Feil, E.J., Enright, M.C. 2004. Analyses of clonality and the evolution of bacterial pathogens. Curr. Opin. Microbiol. **7:** 308-313.

Fitzgerald, C., Sherwood, R., Gheesling, L.L., Brenner, F.W., Fields, P.I. 2003. Molecular analysis of the *rfb* O Antigen gene cluster of *Salmonella enterica* serogroup O:6,14 and development of a serogroup-specific PCR assay. Appl. Environ. Microbiol. **69:** 6099-6105.

Flügge, C. 1886. *B. mycoides* in **'**Die Mikroorganismen', p 324. Vogel, Leipzig.

Gaviria Rivera, A.M., Granum, P.E., Priest, F.G. 2000. Common occurrence of enterotoxin genes and enterotoxicity in *Bacillus thuringiensis*. FEMS Microbiol. Let. **190:** 151-155.

Gaviria Rivera, A.M., Priest, F.G. 2003. Molecular Typing of *Bacillus thuringiensis* serovars by RAPD-PCR. System. Appl. Microbiol. **26:** 254-261.

Gaviria Rivera, A.M., Priest, F.G. 2003. Pulsed field gel electrophoresis of chromosomal DNA reveals a clonal population structure to *Bacillus thuringiensis* that relates in general to crystal protein content. FEMS Microbiol. Lett. **223:** 61-66.

Gillis, M., Vandamme, P., De Vos, P., Swinghs, J., Kersters, K. 2001. Polyphasic Taxonomy. Bergey's manual of Systematic Bacteriology. Springer. **1:** 2nd ed. 43-48.

Gilmore, M.S., Cruz-Rodz, A.L., Leimeister-Wachter, M., Kreft, J., Werner, G. 1989. A *Bacillus cereus* cytolytic determinant, cereolysin AB, which comprises the phospholipaseC and sphingomyelinase genes: nucleotide sequence and genetic linkage. J. Bacteriol. **171:** 744-753.

Gohar, M., Økstad, O.A., Gilois, N., Sanchis, V., Kolstø, A-B., Lereclus, D. 2002. Two-dimensional electrophoresis analysis of the extracellular proteome of *Bacillus cereus* reveals the importance of the PlcR regulon. Proteomics. **2:** 784-791.

Gordon, R.E., Haynes, W.C., Pang, C.H-N. 1973. The Genus *Bacillus*. Agriculture Handbook no 427. Washington D.C: United States Department of Agriculture.

Granum, P.E., 1994. *Bacillus cereus* and its toxins. J. Appl. Bact. Symposium Supplement. **76:** 61S-66S.

Granum, P.E., Lund,T. 1997. Bacillus cereus and its food poisoning toxins. FEMS Mirobiol. Lett. **157:** 223-228.

Granum, P.E., O'Sullivan, K., Lund, T. 1999. The sequence of the non-haemolytic enterotoxin operon from *Bacillus cereus*. FEMS Microbiol. Lett. **177:** 225-229.

Guttmann, D.M., Ellar, D.J. 2000. Phenotypic and genotypic comparisons of 23 strains from the *Bacillus cereus* complex for a selection of known and putative *B. thuringiensis* virulence factors. FEMS Microbiol. Lett. **188:** 7-13.

Hanage, W.P., Fraser, C., Spratt,B. 2005. Fuzzy species among recombinogenic bacteria. BMC Biol. **3:** 1-6.

Harrel, L.J., Andersen, G.L., Wilson, K.H. 1995. Genetic variability of *Bacillus anthracis* and related species. J. Clin. Microbiol. **33:** 1847-1850.

Harwood, C.R. 1989. Taxonomy of *Bacillus*. In: Bacillus (Biotechnology Handbooks 2). Plenum Press. New York. P 11.

Heinrichs, J.H., Beecher, D.J., Macmillan, J.D., Zilinskas, B. 1993. Molecular cloning and characterization of the *hblA* gene encoding the B component of hemolysin Bl from *Bacillus cereus*. J. Bacteriol. **175:** 6760-6766.

Helgason, E., Caugant, D.A., Lecadet, M-M., Chen, Y., Mahillon, J., Lovgren, A., Hegna, I., Kvaloy, K., Kolstø, A.B. 1998. Genetic diversity of *Bacillus cereus / B. thuringiensis* isolates from natural sources. Curr. Microbiol. **37:** 80-87.

Helgason, E., Caugant, D.A., Olsen, I., Kolstø, A.B. 2000 a. Genetic structure of population of *Bacillus cereus* and *B. thuringiensis* isolates associated with periodontis and other human infections. J. Clin. Microbiol. **38:** 1615-1622.

Helgason, E., Okstad O.A., Caugant D. A., Johansen H. A., Fouet A., Mock M.,. Hegna I, and Kolstø, A-B. 2000 b. *Bacillus anthracis, Bacillus cereus* and *Bacillus thuringiensis* - one species on the basis of genetic evidence. Appl. Environ. Microbiol. **66:** 2627-30.

Helgason, E., Tourasse, N.J., Meisal, R., Caugant, D.A., Kolstø, A-B. 2004. A multilocus sequence typing scheme for bacteria of the *Bacillus cereus* group. Appl. Environ. Microbiol. **70:** 191-201.

Hernandez, E., Ramisse, F., Ducoureau, J-P., Cruel, T., Cavallo, J-D. 1998. *Bacillus thuringiensis* subsp. *konkukian* (Serotype H34) superinfection: case report and experimental evidence of pathogenicity in immunosuppressed mice. J. Clin. Microbiol. **36:** 2138-2139.

Hill, K.K., Ticknor,L.O., Okinaka, R.T., Asay, M., Blair, H.,Bliss, K.A., Laker, M., Pardington, P.E., Richardson, A.P., Tonks, M., Beecher, D.J., Kemp, J.D., Kolstø, A-B., Wong, A.C.L., Keim, P., Jackson, P.J. 2004**.** Fluorescent amplified fragment length polymorphism analysis of *Bacillus anthracis, Bacillus cereus* and *Bacillus thuringiensis* isolates. Appl. Environ. Microbiol. **70:** 1068-1080

Hoffmaster, A.R., Ravel, J., Rasko, D.A., Chapman, G.D., Chute, M.D., Marston, C.K., De, B.K., Sacchi, C.T., Fitzgerald, C., Mayer, L.W., Maiden, M.C.J., Priest, F.G., Barker, M., Jiang, L., Cer, R.Z., Rilstone, J., Peterson, S.N., Weyant, R.S., Galloway, D.R., Read, T.D., Popovic, T., Fraser, C.M. 2004. Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. Proc. Natl. Acad. Sci. USA **101:** 8449-8454

Hofte, H., Whiteley, H.R. 1989. Insecticidal crystal proteins of *Bacillus thuringiensis* .Microbiol. Rev. **53:** 242-255.

Holmes, E.C., Urwin, R., Maiden, M.C.J. 1999. The influence of recombination on the population structure and evolution of the human pathogen *Neisseria meningitidis*. Mol. Biol. Evol. **16:** 741-749.

Horwood, P.F., Graham, W., Burgess, H., Oakey, J. 2004. Evidence for non-ribosomal peptide synthetase production of cereulide (the emetic toxin) in *Bacillus cereus.* FEMS Mirobiol. Lett. **236:** 319-324.

Hoton, F.M., Andrup, L., Swiecicka, I., Mahillon. 2005. The cereulide genetic determinants of emetic *Bacillus cereus* are plasmid-borne. Microbiol. **151:** 2121-2124.

Huson D.H. 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics **14:** 68-73.

Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatral, V., Bhattacharyya, A., Reznik, G., Mikhallova, N., Lapidus, A., Chu, L., Mazur, M., Goltsman, E., Larsen, N., D'Souza, M., Walunas, T., Grechkin, Y., Pusch, G., Haselkorn, R., Fonstein, M., Ehrlich, S.D., Overbeek, R., Kyrpides, N. 2003. genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. Nature **423:** 87-91.

Jensen, G. B., Hansen, B. M., Eilenberg, J., Mahillon, J. 2003. The hidden lifestyles of *Bacillus cereus* and relatives. Environ. Microbiol. **5:** 631-640.

Jolley, K.A., Feil, E.J., Chan, M-S., Maiden, M.C.J. 2001. Sequence type analysis and recombinational tests (START). Bioinformatics. **17(12):** 1230-1.

Jolley, K.A., Chan, M-S., Maiden, M.C.J. 2004. mlstdbNet – distributed multi-locus sequence typing (MLST) databases. BMC Bioinformatics **5:** 1-8.

Joung, K.B., Côte, J .C. 2001 a. Phylogenetic analysis of *Bacillus thuringiensis* serovars based on 16S rRNA gene restriction fragment length polymorphisms. J. App. Microbiol. **90:** 115-122.

Joung, K.B., Côte, J .C. 2001 b. A phylogenetic analysis of *Bacillus thuringiensis* serovars by RFLP-based ribotyping. J. App. Microbiol. **91:** 279-289.

Joung, K. B., Cote, J. C. 2002. A single phylogenetic analysis of Bacillus thuringiensis strains and bacilli species inferred from 16S rRNA gene restriction fragment length

polymorphism is congruent with two independent phylogenetic analyses. 2002. J. App. Microbiol. **93:** 1075-1082.

Kaneko, T., Nozaki, R., Aizawa, K. 1978. Deoxyribonucleic acid relatedness between *Bacillus anthracis, Bacillus cereus* and *Bacillus thuringiensis.* Microbiol. Immunol. **22:** 639-641.

Keim, P., Kalif, A., Schupp, J., Hill, K., Travis, S.E., Richmond, K., Adair, D.M., Hugh-Jones, M., Kuske, C.R., Jackson, P. 1997. Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymorphism markers. J. Bacteriol. **179:** 818-824.

Keim, P., Price, L.B., Klevytska, A.M., Smith, K.L., Schupp, J.M., Okinaka, R., Jackson, P.J., Hugh-Jones, M.E. 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J. Bacteriol. **182:** 2928-2936.

Kim K, Seo J, Wheeler K, Park C, Kim D, Park S, Kim W, Chung SI, Leighton T. 2005. Rapid genotypic detection of Bacillus anthracis and the Bacillus cereus group by multiplex real-time PCR melting curve analysis. FEMS Immunol. Med. Microbiol. **43:** 301-10.

Kim, W., Hong, Y., Yoo, J., Lee, W., Choi, C., Chung, S. 2001. Genetic relationships of *Bacillus anthracis* and closely related species based on variable-number tandem repeat analysis and BOX-PCR genomic fingerprinting. FEMS Microbiol. Lett. **207:** 21-27.

King RD, Saqi M, Sayle R and Sternberg MJ. 1997. DSC: public domain protein secondary structure predication. Comput. Appl. Biosci. **13:** 473-474.

Koehler T. M. 2002. *Bacillus anthracis* genetics and virulence gene regulation. Curr. Top. Microbiol. Immunol. **271:** 143-64.

Kotiranta, A., Lounatmaa, K., Haapasalo, M. 2000. Epidemiology and Pathogenesis of *Bacillus cereus* infections. Microbes and Infection 189-198.

La Duc, M.T., Satomi, M., Agata, N., Venkateswaran, K. 2004. *gyrB* as a phylognetic discriminator for members of the *Bacillus anthracis-cereus-thuringiensis* group. J. Microbiol. Meth. **56:** 383-394.

Lan, R., Reeves, P.R. 2000. Intraspecies variation in bacterial genomes: the need for a species genome concept. Trends Microbiol. **8:** 396-401.

Lan, R., Reeves, P.R. 2001. When does a clone deserve a name ? A perspective on bacterial species based on population genetics. Trends Microbiol. **9:** 419-424.

Lan, R., Chehani Alles, M., Donohoe, K., Martinez, M.B., Reeves, P.R. 2004. Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella* spp. Infect Immun. **72:** 5080-5088.

Laskowski, R. European Bioinformatics Institute, Wellcome trust Genome Campus, Hinxton, Cambridge, CB10 1SD

Lecadet, M. M., Chaufaux, J., Ribier, J., Lereclus, D. 1992**.** Construction of novel *Bacillus thuringiensis* strains with different insecticidal activities by transduction and transformation. Appl. Environ. Microbiol. **58:** 840-849.

Lecadet, M.M., Frachon, E., Cosmao Dumanoir, V., Ripouteau, H., Hamon, S., Laurent, P., Thiéry, I. 1999. Updating the H-antigen classification of *Bacillus thuringiensis*. J. Appl. Microbiol**. 86:** 660-672.

Lechner, S., Mayr, R., Francis, K.P., Pruß, B.M., Kaplan, T., Weißner-Gunkel, E., Stewart, G.S.A.B., Scherer, S. 1998. *Bacillus weihenstephanensis* sp. nov is a new psychrotolerant species of the *Bacillus cereus* group. Int. J. System. Bacteriol. **48:** 1373-1382.

Lereclus, D., Agaisse, H., Gominet, M., Salamitou, S., Sanchis, V. 1996. Identification of a *Bacillus thuringiensis* gene that positively regulates transcription of the phosphatidylinositol-specific phospholipase C gene at the onset of the stationary phase. J. Bacteriol. **178:** 2749-2756.

Lindbäck, T., Fagerlund, A., Rødland, M.S., Granum, P.E. 2004. Characterization of the *Bacillus cereus* Nhe enterotoxin. Microbiol. **150:** 3959-3967.

Logan, N. A., Berkeley, R.C.W. 1981. Classification and identification of the genus *Bacillus* using API tests. In 'The Aerobic Endospore-forming bacteria: Classification and Identification'. Pp. 106-140. Ed. R. C. W. Berkeley & M. Goodfellow. London & New York, Academic Press

Lund, T., Granum, P.E. 1997. Comparison of biological effect of the two different enterotoxin complexes isolated from three different strains of *Bacillus cereus*. 1997. Microbiol. **143:** 3329-3336.

Lund, T., Granum, P.E. 1999. The 105-kDa protein component of *Bacillus cereus* on-haemolytic enterotoxin (Nhe) is a metalloprotease with gelatinolytic and collagenolytic activity. FEMS Microbiol. Lett. **178:** 355-361.

Lund, T., Buyser, M-L., Granum, P.E. 2000. A new cytotoxin from *Bacillus cereus* that may cause necrotic enteritis. Mol. Microbiol. **38:** 254-261.

Maiden, M. C. J. 1993. Population genetics of a transformable bacterium: The influence of horizontal genetic exchange on the biology of *Neisseria meningitides***.** FEMS Microbiol. Lett . **112:** 243-250.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M. and Spratt, B.G. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA **95:** 3140-3145

Margulis, L., Jorgensen J. Z., Dolan, S., Kolchinsky, R., Rainey, F. A., Lo, S. C. 1998. The Arthromitus stage of *Bacillus cereus*: intestinal symbionts of insects. Proc. Natl. Acad. Sci. USA **95:**1236-1241.

Maynard Smith, J., Smith, N.H., O'Rourke, M., Spratt, B.G. 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. USA **90:** 4384-4388.

Mishler, B.D., Donoghue, M.J. 1982. Species concepts: a case for pluralism. Syst. Zool. **31:** 491-503.

Myers, L. L., Guinee, P. A. Occurrence and characteristics of enterotoxigenic Escherichia coli isolated from calves with diarrhea. Infect. Immun. **13:**1117-1119.

Nakamura, L. K. 1994. DNA relatedness among *Bacillus thuringiensis* serovars. Int. J. System. Bacteriol. **44:** 125-129.

Nakamura, L.K., Jackson, M.A. 1995. Clarification of the taxonomy of *Bacillus mycoides.* Int. J. System. Microbiol. **45:** 46-49.

Nakamura, L.K. 1998. *Bacillus pseudomycoides* sp.nov. Int. J. Syst. Bacteriol. **48:** 1031-1035.

Noller, A.C., McEllistrem, M.C., Stine, O.C., Morris, J.G. Jnr., Boxrud, D.J., Dixon, B., Harrison, L.H. 2003. Multilocus sequence typing reveals a lack of diversity among *Escherichia coli* 0157:H7 isolates that are distinct by pulsed-field gel electrophoresis. J. Clin. Microbiol. **41:** 675-679.

Okinaka, R.T., Cloud, K., Hampton, O., Hoffnaster, A., Hill, K., Keim, P., Koehler, T., Lamke, G., Kumano, S., Manter, D,. Martinez, Y., Ricke, D., Svensson, R., Jackson, P. 1999. Sequence, assembly and analysis of pXO1 and pXO2. J. App. Bacteriol. **87:** 261-262.

Økstad, A.O., I.H., Lindback, T., Rishovd, A-L., and Kolstø, A.B. 1999. Genome organisation is not conserved between *Bacillus cereus* and *Bacillus subtilis.* Microbiol. **145:** 621-631.

Økstad, A.O., Gominet, M., Purnelle, B., Rose, M., Lereclus, D., Kolstø, A.B. 1999**.** Sequence analysis of three *Bacillus cereus* loci carrying PlcR-regulated genes encoding degradative enzymes and enterotoxin. Microbiol. **145:** 3129-3138.

Page, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences **12**: 357-358.

Park HW, Ge B, Bauer LS, Federici BA. 1998. Optimization of Cry3A yields in *Bacillus thuringiensis* by use of sporulation-dependent promoters in combination with the STAB-SD mRNA sequence. Appl. Environ. Microbiol. **64:** 3932-3938.

Pirttijarvi, T.S.M., Ahonen, L.M., Maunuksela,L.M., Salkinoja-Salonen, M.S., 1998
Bacillus cereus in a whey process. Int. J. Food Micro. **44:** 31-41.

Pirttijarvi, T.S.M., Andersson, M.A., Scoging, A.C., Salkinoja-Salonen, M.S. 1999. Evaluation of Methods for Recognising Strains of the *Bacillus cereus* Group with Food Poisoning Potential Among Industrial and Environmental Contaminants. System. Appl. Microbiol. **22:** 133-144.

Pirttijarvi, T.S.M., Andersson, M.A., Salkinoja-Salonen, M.S. 2000. Properties of *Bacillus cereus* and other bacilli contaminating biomaterial-based industrial processes. Int. J. Food Microbiol. **60:** 231-239.

Pomerantsev, A.P., Kalnin, K.V., Osorio, M., Leppla, S.H. 2003. Phosphatidylcholine-Specific Phosphlipase C and Sphingomyelinase Activities in Bacteria of the *Bacillus cereus* Group. Infect. Immun. **71:** 6591-6606.

Priest, F.G., Dewar, S.J. 2000. Bacteria and Insects in 'Applied Microbial Systematics', pp. 165-202. Ed. F. G. Priest & M. Goodfellow. Kluwer Academic Publishers, The Netherlands.

Priest, F.G., Goodfellow, M., Todd, C. 1988. A numerical classification of the genus *Bacillus.* J. Gen. Microbiol. **134:** 1847-1882.

Priest, F.G., Kaji, D. A., Rosata, Y.B., Canhos, V.P. 1994. Characterization of *Bacillus thuringiensis* and related bacteria by ribosomal RNA gene restriction fragment length polymorphisms. Microbiol. **140:** 1015-1022.

Priest, F.G., Barker, M., Baillie, L.W.J., Holmes, E.D., Maiden, M.C.J. 2004. Population structure and evolution of the *Bacillus cereus* group. J. Bacteriol. **186:** 7959-7970

Pruß, M.B., Dietrich, R., Nibler, B., Martlbauer, E., and Scherer, S. 1999. The hemolytic enterotoxin HBL is broadly distributed among species of the *Bacillus cereus* group. Appl. Environ. Microbiol. **65:** 5436 - 5442.

Rasko, D.A., Ravel, J., Økstad, A.O.,Helgason, E., Cer, R.Z., Jiang, L., Shores, K.A., Fouts, D.E., Tourasse, N.J., Angiuoli, S.V., Kolonay, J., Nelson, W.C., Kolstø, A.B., Fraser, C.M., Read, T.D. 2004. The genome sequence of *Bacillus cereus* ATCC 10987 reveals metabolic adaptations and a large plasmid related to *Bacillus anthracis* pXO1. Nucl. Acids Res. **32:** 977-988.

Rasko, D.A., Altherr, M.R., Han, C.S., Ravel, J. Genomics of the Bacillus cereus group of organisms. 2005. FEMS Microbiol. Rev. **29:** 303-329.

Read T,D., Peterson S,N., Tourasse N, Baillie L.W., Paulsen I.T., Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R., Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M, Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M., DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H., Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D., Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F., Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C., Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L., Thomason B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.B., Fraser C.M. 2003 The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. Nature **423:**81-86.

Ripabelli, G., McLauchlin, J., Mithani, V., Threlfall, E. J. 2000. Epidemological typing of *Bacillus cereus* by amplified fragment length polymorphism. Lett. Appl. Microbiol. **30:** 358-363.

Roselló-Mora, R. 2003. Opinion: The species problem, can we achieve a universal concept ? System. Appl. Microbiol. **26:** 323-326.

Rozas, J., Rozas. R. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics. **15:** 174-175.

Ryan, P.A., Macmillan, J.D., Zilinskas, B.A. 1997. Molecular cloning and characterization of the genes encoding the L_1 and L_2 components of hemolysin BL from *Bacillus cereus*. J. Bacteriol. **179:** 2551-2556.

Saitou, N. and Nei, M. 1987**.** The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. **4:** 406-25.

Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D.R., Dean, D.H. 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. **62:** 775-806.

Scobie, H. M., Young, J.A. 2005. Interactions between anthrax toxin receptors and protective antigen. Curr. Opin. Microbiol. **8:** 106-12.

Shisa, N., Wasano, N., Ohgushi, A., Lee, D-H., Ohba, M. 2002. Extremely high frequency of common flagellar antigens between *Bacillus thuringiensis* and *Bacillus cereus.* FEMS Microbol. Lett. **213:** 93-96.

Skelton, P. 1993. Species, speciation and extinction in 'Evolution, A Biological and Palaeontological Approach'. pp 372-380. The Open University. Text for course S365.

Skerman, V. B. D, McGowan, V., Sneath, P. H. A., 1980. Approved lists of bacterial names. Int. J. Syst. Bacteriol. **30:** 225-420.

Slamti, L., Perchat, S., Gominet, M., Vilas-Boas, G., Fouet, A., Mock, M., Sanchis, V., Chaufaux, J., Gohar, M., Lereclus, D. 2004. Distinct mutations in PlcR explain why some strains of the *Bacillus cereus* group are nonhemolytic. J Bacteriol. **186:** 3531-

3538.

Smith, N. R., Gordon, R. E., Clark, F. E., 1952. Aerobic Sporeforming Bacteria. US Department of Agriculture. Agriculture Monograph No**. 16**: Washington D.C. US Government Printing Office.

Sneath, P.H.A., Sokal, R.R. 1973. Numerical Taxonomy: the principles and practice of numerical classification. San Francisco: Freeman.

Sneath, P., 1986. Endospore-forming gram-positive rods and cocci. In Bergey's Manual of Systematic Bacteriology, 1st ed. Section 13, p. 1104-1105. Williams and Wilkins, USA.

Somerville, H. J., Jones, M. L. 1972. DNA competition studies within *Bacillus cereus* group of bacilli. J. Gen. Microbiol. **73:** 257-265.

Spratt, B. G. 1999. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the Internet. Curr. Opinion Microbiol. **2:** 312-316

Spratt, B. G., Hanage, W.P., Feil, E.J. 2001. The relative contributions of recombination and point mutation to the diversification of bacterial clones. Curr. Opinion Microbiol. **4:** 602-606.

Spratt, B. G., Hanage, W. P., Bao, L., Aanensen, D. M., Feil, E. J. 2004. Displaying the relatedness among isolates of bacterial species – the eBURST approach. FEMS Microbiol Lett. **241:** 129-134..

Stanier, R.Y., Adelberg, E.A., Ingrahan, J.L. 1977. The enteric group and related organisms in 'General Microbiology', pp. 618-623. Prentice-Hall Inc. USA.

Stenfors, L.P., Mayr, R., Scherer, S., Granum, P.E. 2002. Pathogenic potential of fifty *Bacillus weihenstephansis* strains. FEMS Microbiol. Lett. **215:** 47-51.

Stryer, L. 1995 a. Membrane Structure and Dynamics in 'Biochemistry'. pp. 263-289 W.H. Freeman and Company. NY.

Stryer, L. 1995 b. Citric Acid Cycle in 'Biochemistry'. pp. 340-345. W.H. Freeman and Company. NY.

Templeton, E.R. 1989. The meaning of species and speciation: a genetic perspective. In 'Speciation and Its Consequences'. Sunderland, M.A: Sinauer. 3-27

Thomas, W.E., Ellar, D.J. 1983**.** *Bacillus thuringiensis var israelensis* crystal delta-endotoxin: effects on insect and mammalian cells in vitro and in vivo**.** J. Cell Sci. **60:** 181-197.

Thompson J.D., Higgins D.G., Gibson T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673-4680.VIEW

Ticknor, L. O., Kolstø, A. B., Hill, K.K., Keim, P., Laker, M.T., Tonks, M., Jackson, P.J. 2001. Fluorescent Amplified Fragment Length Polymorphism Analysis of Norwegian *Bacillus cereus* and *Bacillus thuringiensis* Soil Isolates. Appl. Environ. Microbiol. **67:** 4863-4873

Turnbull, P.C.B., Jørgensen, K., Kramer, J.M., Gilbert, R.J., Parry, J.M. 1979. Severe clinical conditions associated with *Bacillus cereus* and the apparent involvement of exotoxins. J. Clin. Path. **32:** 289-293.

Turnbull, P.C. 2002. Introduction: anthrax history, disease and ecology. Review, ed. Koehler T.M. Curr. Top. Microbiol. Immunol. **271:** 1-19.

Urwin, R., Maiden, M.C. 2003. Multi-locus sequence typing: a tool for global epidemiology. Trends Microbiol. **11:** 479-487

Vilas-Boas, G., Sanchis, V., Lereclus, D., Lemos, M.V.F., Bourguet D. 2002. Genetic differentiation between sympatric populations of *Bacillus cereus* and *Bacillus thuringiensis*. Appl. Environ. Microbiol. **68:** 1414-1424.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Frijters, A., Pot, J., Peleman, J., Kuiper, M., Zabeau, M. 1995. AFLP: a new technique for DNA fingerprinting.Nucleic Acids Res. **23:** 4407-4414.

Yamada, S., Ohashi, E., Agata, N., Venkateswaran, K. 1999. Cloning and nucleotide sequence analysis of *gyrB* of *Bacillus cereus, B. thuringiensis, B. mycoides* and *B. anthracis* and their application to the detection of *B. cereus* in rice. App. Environ. Microbiol. **65:** 1483-1490.

FEMS Microbiology Letters 245 (2005) 179–184

FEMS MICROBIOLOGY Letters

www.fems-microbiology.org

Multilocus sequence typing reveals that *Bacillus cereus* strains isolated from clinical infections have distinct phylogenetic origins

Margaret Barker [a], Bishan Thakker [b], Fergus G. Priest [a,*]

[a] *School of Life Sciences, Heriot Watt University, Edinburgh, Edinburgh EH14 4AS, UK*
[b] *Department of Medical Microbiology, Glasgow Royal Infirmary, Glasgow, UK*

Received 26 January 2005; received in revised form 3 March 2005; accepted 3 March 2005

First published online 19 March 2005

Edited by A. Klier

Abstract

Eight strains of *Bacillus cereus* isolated from bacteremia and soft tissue infections were assigned to seven sequence types (STs) by multilocus sequence typing (MLST). Two strains from different locations had identical STs. The concatenated sequences of the seven STs were aligned with 65 concatenated sequences from reference STs and a neighbor-joining tree was constructed. Two strains were distantly related to all reference STs. Three strains were recovered in a clade that included *Bacillus anthracis*, *B. cereus* and rare *Bacillus thuringiensis* strains while the other three strains were assigned to two STs that were more closely affiliated to most of the *B. thuringiensis* STs. We conclude that invasive *B. cereus* strains do not form a single clone or clonal complex of highly virulent strains.

Keywords: *Bacillus anthracis*; *Bacillus cereus*; *Bacillus thuringiensis*; MLST; Clone; Phylogeny

1. Introduction

Bacillus cereus is commonly found in the soil and associated with plants [1]. It is best known for causing two forms of mild food poisoning characterized by predominantly diarrhoea or vomiting [2,3]. Occasionally *B. cereus* is implicated in more serious invasive infections. For example, it is responsible for a particularly severe form of endophthalmitis which may result in loss of functional vision or even blindness [4], and bacteremia in the immunocompromised host [5] as well as in preterm neonates [6]. Meningitis, pneumonia, urinary tract infections and fatal fulminant liver failure have also been attributed to *B. cereus* infections (reviewed in [3]). *B. cereus* strains secrete a battery of extracellular enzymes and toxins including phospholipases C, sphingomyelinase and various hemolysins that contribute to their virulence [7,8].

B. cereus is the parent species of a wider group of bacteria encompassing several additional species, *Bacillus anthracis*, *Bacillus mycoides*, *Bacillus pseudomycoides*, *Bacillus thuringiensis* and *Bacillus weihenstephanensis* that are so closely related to *B. cereus* that they are generally considered a single taxospecies [1,9]. Plasmid encoded virulence factors distinguish *B. anthracis*, the causative agent of anthrax [10], and *B. thuringiensis*, a common insect pathogen that is used as a commercial bioinsecticide [11]. *B. mycoides* and *B. pseudomycoides* strains are differentiated from *B. cereus* by their distinctive rhizoid colony morphology [12] and *B. weihenstephanensis* strains are relatively psychrotolerant [13]. Accordingly, the *B. cereus* group can be treated as a

* Corresponding author. Tel.: +44 131 451 3464; fax: +44 131 451 3009.
E-mail address: f.g.priest@hw.ac.uk (F.G. Priest).

0378-1097/$22.00
doi:10.1016/j.femsle.2005.03.003

single species for population analysis and has been subject to both multilocus enzyme electrophoresis (MEE) [14,15] and multilocus sequence typing (MLST) [16,17] analyses that have indicated a reasonably strong clonal structure to the organisms with evidence for specific clones associated with the emetic form of food poisoning [18], periodontal disease [14] and some serovars of *B. thuringiensis* with particular entomopathogenic traits [19].

In this study we have examined eight strains of *B. cereus* isolated from cases of bacteremia and from soft tissue infections by MLST to assess the phylogenetic origins of the strains. We show that two of the strains were identical for the seven partial gene sequences that we analyzed, while the other strains were phylogenetically distinct. We conclude that strains of *B. cereus* involved in opportunistic infections do not belong to a single clonal complex.

2. Materials and methods

2.1. Strains and molecular methods

Five strains of *B. cereus* were received from J. McLauchlin (Food Safety Microbiology Laboratory, Health Protection Agency, London) that had been associated with bacteremia (Table 1). Three strains were isolated from soft tissue infections and identified as *B. cereus* following growth on *B. cereus* selective agar (Oxoid) and phenotypic inspection (Table 1). Methods for strain preservation, bacterial growth, isolation of DNA have been described previously [17]. We amplified by PCR fragments of seven housekeeping genes (*glpF*, *gmk*, *ilvD*, *pta*, *pur*, *pycA* and *tpi*) as described on the website for MLST of *B. cereus* (www.pubmlst.org/bcereus) using the standard primers with primer option 1 for the *ilvD* gene. PCR fragments were sequenced in both directions using the amplification primers to provide unambiguous sequence data.

2.2. Data analysis

The allele sequences were compared with existing allele sequences using the *B. cereus* MLST website and given new allele numbers if they differed from known alleles. The seven allele numbers define a sequence type (ST) a number for which was assigned to each new allele combination. The concatenated sequences for the seven gene fragments for all sequence types were constructed and downloaded using the *B. cereus* MLST website. Multiple alignment of all the concatenated sequences was carried out using CLUSTAL W [20]. Neighbor-joining trees were derived from the alignments using the tree building facility in CLUSTAL W and visualized using TreeView (www.taxonomy.zoology.gla.ac.uk/rod/treeview.html). Clonal groups or lineages were further analyzed using SplitsTree [21]. The position and frequency of polymorphisms in the allele nucleotide sequences were detected using the START program [22].

3. Results

3.1. Phylogenetic origins of the invasive strains

All eight clinical strains of *B. cereus* were isolated in 2003. Five strains were derived from blood of patients of varying age and health and the remaining three from tissue infections (Table 1). The last three cases responded successfully to ciprofloxacin treatment. Most of the clinical isolates had unique combinations of alleles providing STs that were new to the database, the exception was strain 172560W which shared the same ST as a strain of *B. thuringiensis* serovar pakistani isolated in Chile. Strains R2955/03 and R3149/03 had identical STs despite their origins from different hospitals (Table 1).

We constructed a neighbor-joining tree from the concatenated sequences of the seven alleles for the clinical isolates together with concatenated sequence from selected STs from the database (Fig. 1). Two of the isolates, R3238/03 (ST-72) and R3098/03 (ST-74), were distantly related to all other STs and formed independent lines of descent. Indeed ST-72 and ST-74 each had unique alleles which did not occur in any other ST in the database (the allele numbers and sequences for all STs are available from www.pubmlst.org/bcereus).

Table 1
Origins of *Bacillus cereus* isolates used in this study

Strain	Source	Patient/age	Predisposing factors	Location in UK	Sequence type (ST)
R2955/03	Blood	Male, 50	Unknown	Bristol	73
R3039/03	Blood	Male, 70	Unknown	Leicester	75
R3098/03	Blood	Male, 90	Post blood transfusion	Hull	74
R3149/03	Blood	Female, 86	Myeloma	Manchester	73
R3238/03	Blood	Female, 65	Unknown	Oxford	72
168287M	Hematoma	Male, 66	Post-operative infection following prosthetic elbow joint surgery	Glasgow	77
191560K	Calf tissue	Female, 58	Insect bite	Glasgow	76
172560W	Burn wound	Female, 49	20% full thickness burns to legs	Glasgow	18

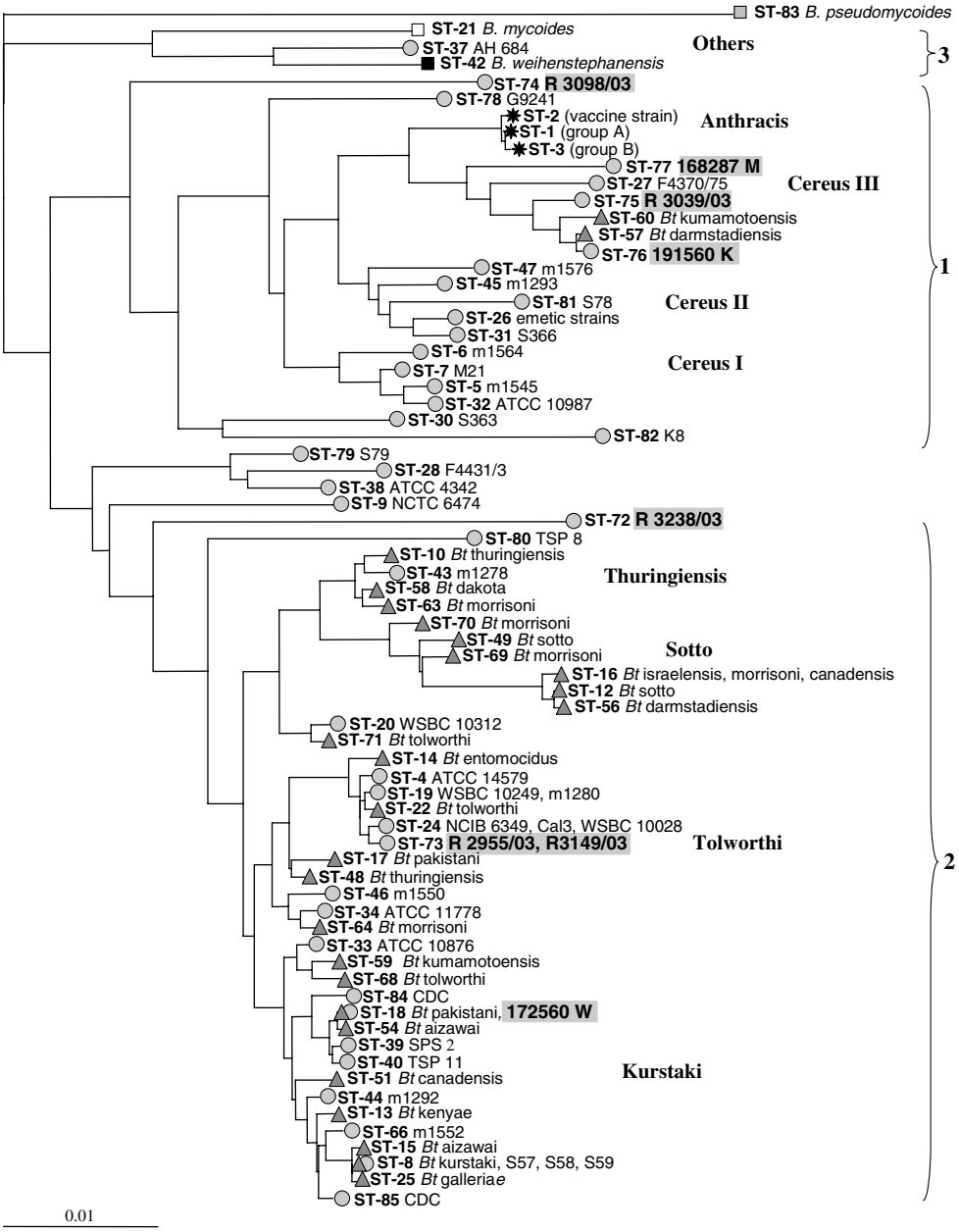
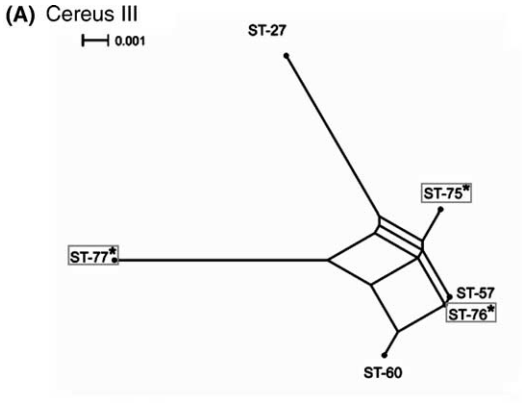


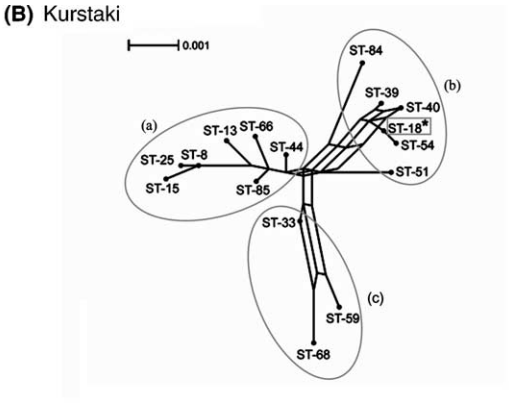
Fig. 1. Neighbor-joining tree showing the relationships of the clinical isolates of *Bacillus cereus* (shaded strain numbers) with reference strains of the *B. cereus* group. The tree has been rooted with *Bacillus pseudomycoides* (ST-83). Clades are indicated by numbers and lineages by names (see text and [17] for details). Bar is substitutions/site. Species designations for strains are: star, *B. anthracis*; shaded circle, *B. cereus*; open square, *B. mycoides*; shaded square, *B. pseudomycoides*; shaded triangle, *B. thuringiensis*; filled square, *B. weihenstephanensis*.

The *B. anthracis–B. cereus–B. thuringiensis* population comprises two major clades: clade 1 which encompasses *B. anthracis*, numerous *B. cereus* and rare *B. thuringiensis* strains, and clade 2 which comprises *B. cereus* and *B. thuringiensis* strains (Fig. 1). A third clade of mixed organisms has been labeled as ''others'' in Fig. 1. Within these clades various lineages have been defined [21] and most of the clinical *B. cereus* isolates were associated with these lineages.

One strain from septicemia (R 3039/03) and two soft tissue isolates (191560K and 168297M) were recovered in lineage Cereus III, a member of clade 1, as STs 75, 76 and 77, respectively (Table 1). A splitsgraph of this cluster revealed the close phylogenetic origins of STs 57, 60, and 76 and the more loosely affiliated STs 27, 60 and 77 (Fig. 2A). There was considerable variation among alleles in this lineage but it was notable that strain R3039/03, shared the *gmk1* and *pta1* alleles with *B. anthracis* strains.

Of the remaining three strains, two identical isolates from blood (strains R2955/03 and R3149/03) and one from a soft tissue infection (172560M) were assigned



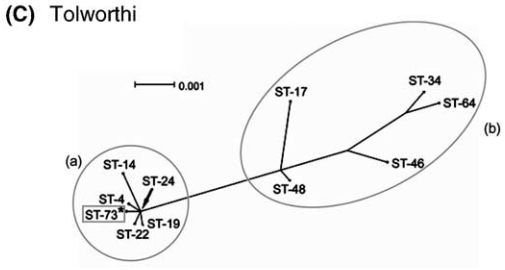


Number of nucleotide changes in sub-lineage Cereus III

ST change	*glpF*	*gmk*	*ilvD*	*pta*	*pur*	*pycA*	*tpi*
ST 57 to ST 27		1	24	6	1	2	
ST 57 to ST 60	2			1			8
ST 57 to ST 75	2	1	4	2	1	2	
ST 57 to ST 76				1			1
ST 57 to ST 77	4	2	15	9	5	2	7

Number of nucleotide changes in sub-lineage Kurstaki (b)

ST change	*glpF*	*gmk*	*ilvD*	*pta*	*pur*	*pycA*	*tpi*
ST 18 to ST 39	1		1		1	1	
ST 18 to ST 40	2				1		
ST 18 to ST 51	2			1	9	1	
ST 18 to ST 54				1			
ST 18 to ST 84	1		2	1	1	6	



Number of nucleotide changes in sub-lineage Tolworthi (a)

ST change	*glpF*	*gmk*	*ilvD*	*pta*	*pur*	*pycA*	*tpi*
ST 24 to ST 4	2		1	2			1
ST 24 to ST 14	1			1	2	5	
ST 24 to ST 19				2			2
ST 24 to ST 22			1	2	1		1
ST 24 to ST 73	2				1		1

Fig. 2. Split decomposition analyses of lineages containing clinical isolates of *B. cereus* (STs boxed) derived from concatenated sequences. (A) lineage Cereus III; (B) lineage Kurstaki; (C) lineage Tolworthi. Sub-lineages are circled and identified by lower case letters in brackets. Associated tables indicate numbers of nucleotide changes between alleles. See text for details.

to separate lineages in clade 2 which comprises a mixed population of *B. cereus* and *B. thuringiensis* strains (Fig. 1). Strain 172560M was recovered in lineage Kurstaki. Split decomposition of the STs in this lineage indicated three principal groups or sub-lineages, designated (a), (b) and (c) in Fig. 2B with the pathogen assigned to sub-lineage (b). This sub-lineage was defined by alleles *gmk9* and *tpi7* with variation in the other alleles.

Two identical isolates from separate cases of bacteremia were assigned to ST-73 in lineage Tolworthi. Split decomposition of this lineage indicated two groups of STs that we have labeled (a) and (b) in Fig. 2C with the clinical isolates located in sub-lineage (a). To determine further the structure and origins of these sub-lineages we undertook an evaluation of the roles of recombination versus mutation in their evolution.

3.2. Mutation versus recombination in the evolutionary descent of B. cereus strains from clinical sources

If we assume an ancestral ST for a group or sub-lineage based on the maximum number of shared alleles within the group, it is possible to compare variant (derived) STs with the assumed ancestor [23]. Single base differences between two alleles are likely to be due to mutation, double changes are less likely to be due to mutation, and when alleles differ at multiple sites it is more likely that the observed variation is due to lateral gene transfer, especially if the variant allele occurs in a non-related ST elsewhere in the dataset [23]. We have used these criteria to indicate the relative roles of recombination and mutation in the *B. cereus* group (Fig. 2).

In lineage Cereus III, ST-57 was designated the ancestor and while single nucleotide changes are extant,

there is strong evidence for recombinational exchange especially in *ilvD*, *pta* and *tpi*. In sub-lineage Kurstaki (b), many variant alleles differed by a single nucleotide change (Fig. 2B), The replacement of *pur12* in ST-18 by *pur28* in ST-51, on the other hand, involved nine changes to an allele which suggests recombinational exchange between members of the sub-lineage. Nevertheless, the evolutionary trend in sub-lineage Kurstaki (b) appeared to be biased towards the mutational. Finally, two blood isolates were assigned to sub-lineage Tolworthi (a). The clone ST-24 is the likely ancestor of this group as indicated by split decomposition and shared alleles (Fig. 2C). The star-like nature of the splits graph, and preponderance of single and double mutational changes suggest predominance of mutation over recombination in this group.

4. Discussion

Opportunistic infections typically occur when an otherwise harmless organism is introduced into a body tissue in sufficient numbers that the organisms can proliferate. *B. cereus* has an array of toxins which enable it to colonize the host and it was possible that clones or clonal complexes with a particular array of virulence factors may be responsible for invasive infections. For example, there is evidence that isolates of *B. cereus* causing periodontal disease are associated with a particular clonal group [14] and the ST-26 strains which cause the emetic form of food poisoning are strongly clonal [18]. However, the results from this study suggest that the ability of *B. cereus* strains to cause invasive disease is not restricted to a particular clonal group of strains, but is a feature of a genomically diverse group of organisms. This lack of correlation between virulence potential and MLST genotype has also been noted in *Staphylococcus aureus* [23] and contrasts with community acquired pathogens such as *Neisseria meningitidis* [24] and *Streptococcus pneumoniae* [25] in which certain clones are particularly invasive. However, we emphasize that this conclusion is based on a small sampling of strains from clinical sources and, it is possible that when more strains have been analyzed, virulent clones or clonal complexes will emerge such as ST-73 in lineage Tolworthi.

The *B. anthracis–B. cereus–B. thuringiensis* population is evolving in two major clades: clade 1 comprising *B. anthracis*, *B. cereus* and rare *B. thuringiensis* strains, and clade 2 encompassing a mixed population of *B. cereus* and *B. thuringiensis* strains [17]. The recovery of three of the clinical *B. cereus* isolates in clade 1, closely related to *B. anthracis* and, in the case of strain R3039/03, sharing *gmk1* and *pta1* alleles with *B. anthracis* strains, is consistent with this part of the population encompassing virulent strains. *B. cereus* G9241 which was associated with an illness resembling inhalation anthrax and contains the anthrax toxin genes but lacks the polyglutamate capsule [26] is also loosely related to clade 1 (Fig. 1). It was therefore a little surprising to find clinical isolates also in clade 2 which is generally associated with *B. thuringiensis* strains which, with rare exceptions are considered harmless [27]. *B. cereus* 172560W from a burns patient was identical to a strain of *B. thuringeinsis* serovar pakistani. We have examined strain 172560W microscopically and it does not make a visibly detectable crystal protein, therefore by definition is not a strain of *B. thuringiensis*.

Evidence for lateral gene transfer among *B. cereus* and *B. thuringiensis* populations has been provided by MEE studies [15]. Here we indicated the relative contributions of mutation and recombination in the evolution of these bacteria by scrutinizing the distribution of nucleotide changes in alleles among closely related strains. Our data suggest that recombination and mutation both contribute to the evolution of the population with predominance of mutation in sub-lineages Kurstaki (b) and Tolworthi (a). Similar analyses indicate that recombinational exchange is also less common than mutation in the related, low G+C organism, *S. aureus* [23].

B. cereus has a history of association with invasive infection [28]. These examples of strains isolated in 2003 indicate that it continues to pose a threat to both healthy and immuncompromised individuals. Here we have established that the bacteria responsible for these infections are not restricted to a single clonal group or lineage within *B. cereus* but are genomically diverse.

Acknowledgments

We thank J. McLauchlin (Health Protection Agency, London) for provision of strains and K. Jolley (University of Oxford) for establishment and assistance with maintenance of the *B. cereus* MLST website. Helen Agnew and Allan Angus assisted with DNA sequencing.

References

[1] Priest, F.G. (1993) Systematics and Ecology of *Bacillus* In: Bacillus Subtilis and Other Gram-positive Bacteria; Biochemistry, Physiology, and Molecular Genetics (Sonenshein, A.L., Hoch, J.A. and Losick, R, Eds.), pp. 3–16. American Society for Microbiology, Washington, DC.

[2] Granum, P.E. and Lund, T. (1997) *Bacillus cereus* and its food poisoning toxins. FEMS Microbiol Lett. 157, 223–228.

[3] Kotiranta, A., Lounatmaa, K. and Haaspasalo, M. (2000) Epidemiology and pathogenesis of *Bacillus cereus* infections. Microbes Infect. 2, 189–198.

[4] Callegan, M.C., Kane, S.T., Cochran, D.C. and Gilmore, M.S. (2002) Molecular mechanisms of *Bacillus* endophthalmitis pathogenesis. DNA Cell Biol. 21, 367–373.

[5] Ginsburg, A.S., Salazar, L.G., True, L.D. and Disis, M.L. (2003) Fatal *Bacillus cereus* sepsis following resolving netropenic entero-

colitis during treatment of acute leukemia. Am. J. Hematol. 72, 204–208.
[6] Hilliard, N.J., Schelonka, R.L. and Waites, K.B. (2003) *Bacillus cereus* bacteremia in a preterm neonate. J. Clin. Microbiol. 41, 3441–3444.
[7] Guttmann, D.M. and Ellar, D.J. (2000) Phenotypic and genotypic comparisons of 23 strains from the *Bacillus cereus* complex for a selection of known and putative *B. thuringiensis* virulence factors. FEMS Microbiol Lett. 188, 7–13.
[8] Pomerantsev, A.P., Kalnin, K.V., Osorio, M. and Leppla, S.H. (2003) Phosphatidylcholine-specific phospholipase C and sphingomyelinase activities in bacteria of the *Bacillus cereus* group. Infect. Immun. 71, 6591–6606.
[9] Helgason, E., Økstad, O.A., Caugant, D.A., Johansen, H.A., Fouet, A., Mock, M., Hegna, I. and Kolstø, A.-B. (2000) *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* – one species on the basis of genetic evidence. Appl. Environ. Microbiol. 66, 2629–2630.
[10] Baillie, L.W. and Read, T.D. (2001) *Bacillus anthracis*, a bug with attitude. Curr. Opin. Microbiol. 4, 78–81.
[11] Aronson, A. (2002) Sporulation and delta-endotoxin synthesis by *Bacillus thuringiensis*. Cell Mol. Life Sci. 59, 417–425.
[12] Nakamura, L.K. and Jackson, M.A. (1995) Clarification of the taxonomy of *Bacillus mycoides*. Int. J. Syst. Bacteriol. 45, 46–49.
[13] Lechner, S., Mayr, R., Francis, K.P., Pruss, B.M., Kaplan, T., Wiessner-Gunkel, E., Stewart, G.S. and Scherer, S. (1998) *Bacillus weihenstephanensis* sp. nov. is a new psychrotolerant species of the *Bacillus cereus* group. Int. J. Syst. Bacteriol. 48, 1373–1382.
[14] Helgason, E., Caugant, D.A., Olsen, I. and Kolstø, A.-B. (2000) Genetic structure of population of *Bacillus cereus* and *B. thuringiensis* isolates associated with periodontitis and other human infections. J. Clin. Microbiol. 38, 1615–1622.
[15] Vilas-Boas, G., Sanchis, V., Lereclus, D., Lemos, M.F.L. and Bourguet, D. (2002) Genetic differentiation between sympatric populations of *Bacillus cereus* and *Bacillus thuringiensis*. Appl. Environ. Microbiol. 68, 1414–1424.
[16] Helgason, E., Tourasse, N.J., Meisal, R., Caugant, D.A. and Kolstø, A.-B. (2004) Multilocus sequence typing scheme for bacteria of the *Bacillus cereus* group. Appl. Environ. Microbiol. 70, 191–201.
[17] Priest, F.G., Barker, M., Baillie, L.W.J., Holmes, E.C. and Maiden, M.C.J. (2004) Population structure and evolution of the *Bacillus cereus* group. J. Bacteriol. 186, 7959–7970.
[18] Ehling-Schulz, M., Svensson, B., Guinebretiere, M.-H., Lindbäck, T., Andersson, M., Schulz, A., Fricker, M., Christiansson, A., Granum, P.E.., Märtlbauer, E., Nguyen-The, C., Salkinoja-salonen, M. and Scherer, S. (2005) Emetic toxin formation of *Bacillus cereus* is restricted to a single lineage. Microbiology 151, 183–197.
[19] Gaviria Rivera, A. and Priest, F.G. (2003) Pulsed field gel electrophoresis of chromosomal DNA reveals a clonal population structure to *Bacillus thuringiensis* that relates in general to crytsal protein content. FEMS Microbiol. Lett. 223, 61–66.
[20] Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J. and Higgins, D.G. (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31, 3497–3500.
[21] Huson, D.H. (1998) SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14, 68–73.
[22] Jolley, K.A., Feil, E.J., Chan, M.S. and Maiden, M.C. (2001) Sequence type analysis and recombinational tests [START]. Bioinformatics 17, 1230–1231.
[23] Feil, E.J., Cooper, J.E., Grundmann, H., Robinson, D.A., Enright, M.C., Berendt, T., Peacock, S.J., Smith, J.M., Murphy, M., Spratt, B.G., Moore, C.E. and Day, N.P.J. (2003) How clonal is *Staphylococcus aureus*? J. Bacteriol. 185, 3307–3316.
[24] Feavers, I.M., Gray, S.J., Urwin, R., Russell, J.E., Bygraves, J.A., Kaczmarski, E.B. and Maiden, M.C.J. (1999) Multilocus sequence typing and antigen sequencing in the investigation of a meningococcal disease outbreak. J. Clin. Microbiol. 37, 3883–3887.
[25] Brueggemann, A.B., Griffiths, D.T., Meats, E., Peto, T., Crook, D.W. and Spratt, B.G. (2003) Clonal relationships between invasive and carriage *Streptococcus pneumoniae* and serotype- and clone-specific differences in invasive disease potential. J. Infect. Dis. 187, 1424–1432.
[26] Hoffmaster, A.R., Ravel, J., Rasko, D.A., Chapman, G.D., Chute, M.D., Marston, C.K., De, B.K., Sacchi, C.T., Fitzgerald, C., Mayer, L.W., Maiden, M.C., Priest, F.G., Barker, M., Jiang, L., Cer, R.Z., Rilstone, J., Peterson, S.N., Weyant, R.S., Galloway, D.R., Read, T.D., Popovic, T. and Fraser, C.M. (2004) Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax. Proc. Natl. Acad. Sci. USA 101, 8449–8454.
[27] Hernandez, E., Ramisse, F., Cruel, T., le Vagueresse, R. and Cavallo, J.D. (1999) *Bacillus thuringiensis* serotype H34 isolated from human and insecticidal strains serotypes H3a3b and H14 can lead to death of immunocompetent mice after pulmonary infection. FEMS Immunol. Med. Microbiol. 24, 43–47.
[28] Turnbull, P.C.B., Jorgensen, K., Kramer, J.M., Gilbert, R.J. and Parry, J.M. (1979) Severe clinical conditions associated with *Bacillus cereus* and the apparent involvement of exotoxins. J. Clin. Pathol. 32, 293–299.

JOURNAL OF BACTERIOLOGY, Dec. 2004, p. 7959–7970 Vol. 186, No. 23
0021-9193/04/$08.00+0 DOI: 10.1128/JB.186.23.7959–7970.2004

Population Structure and Evolution of the *Bacillus cereus* Group†

Fergus G. Priest,[1]* Margaret Barker,[1] Les W. J. Baillie,[2] Edward C. Holmes,[3] and Martin C. J. Maiden[3,4]

School of Life Sciences, Heriot Watt University, Edinburgh,[1] and Department of Zoology[3] and Peter Medawar Building for Pathogen Research,[4] University of Oxford, Oxford, United Kingdom, and Biological Defense Research Directorate, United States Navy Medical Research Center, Silver Spring, Maryland[2]

Received 5 May 2004/Accepted 2 September 2004

Representative strains of the *Bacillus cereus* group of bacteria, including *Bacillus anthracis* (11 isolates), *B. cereus* (38 isolates), *Bacillus mycoides* (1 isolate), *Bacillus thuringiensis* (53 isolates from 17 serovars), and *Bacillus weihenstephanensis* (2 isolates) were assigned to 59 sequence types (STs) derived from the nucleotide sequences of seven alleles, *glpF*, *gmk*, *ilvD*, *pta*, *pur*, *pycA*, and *tpi*. Comparisons of the maximum likelihood (ML) tree of the concatenated sequences with individual gene trees showed more congruence than expected by chance, indicating a generally clonal structure to the population. The STs followed two major lines of descent. Clade 1 comprised *B. anthracis* strains, numerous *B. cereus* strains, and rare *B. thuringiensis* strains, while clade 2 included the majority of the *B. thuringiensis* strains together with some *B. cereus* strains. Other species were allocated to a third, heterogeneous clade. The ML trees and split decomposition analysis were used to assign STs to eight lineages within clades 1 and 2. These lineages were defined by bootstrap analysis and by a preponderance of fixed differences over shared polymorphisms among the STs. Lineages were named with reference to existing designations: Anthracis, Cereus I, Cereus II, Cereus III, Kurstaki, Sotto, Thuringiensis, and Tolworthi. Strains from some *B. thuringiensis* serovars were wholly or largely assigned to a single ST, for example, serovar aizawai isolates were assigned to ST-15, serovar kenyae isolates were assigned to ST-13, and serovar tolworthi isolates were assigned to ST-23, while other serovars, such as serovar canadensis, were genetically heterogeneous. We suggest a revision of the nomenclature in which the lineage and clone are recognized through name and ST designations in accordance with the clonal structure of the population.

The *Bacillus cereus* group comprises closely related gram-positive bacteria that exhibit highly divergent pathogenic properties. Many bacteria classified as *B. cereus* are widely distributed in the environment, with probable reservoirs in the soil (57), and as commensal inhabitants of the intestines of insects (35). Occasionally they are associated with food poisoning (16) and with soft tissue infections, particularly of the eye (9). Other members of the group that are currently classified as *Bacillus thuringiensis* are primarily insect pathogens. These bacteria produce toxins in the form of parasporal crystal proteins that have been widely used for the biocontrol of insect pests (49). Occasionally, *B. thuringiensis* strains are responsible for human infections similar to those caused by strains of *B. cereus* (7, 25). A third pathogenic phenotype is exhibited by *Bacillus anthracis*, a pathogen of mammals and especially ungulates that can cause human disease (36). The principal virulence factors of *B. anthracis* are encoded by genes located on two plasmids: the tripartite toxin genes *pag*, *lef*, and *cya* are carried on plasmid pXO1, while the genes encoding the biosynthesis of the poly-D-glutamate capsule, *capA*, *capB*, and *capC*, are carried on a smaller plasmid, pXO2 (38). Similarly, the crystal protein genes responsible for the major features of insect toxicity of *B. thuringiensis* isolates are almost invariably plasmid encoded (49). The virulence genes of *B. cereus*, on the other hand, are chromosomal (17, 24, 43).

These three species of the *B. cereus* group were first described around the turn of the 19th century, yet despite this long history the relationships between these organisms have yet to be completely resolved (44, 46). Whole-genome DNA hybridization has been unhelpful (28, 37, 50), while conventional markers of chromosomal diversity, such as 16S and 23S rRNA genes, are essentially identical (2, 3). Comprehensive studies using a diverse range of techniques, including genomic mapping (5), pulsed-field gel electrophoresis of chromosomal DNA (4), multilocus enzyme electrophoresis (18, 19), variable number tandem repeat mapping, BOX-PCR fingerprinting (31), amplified fragment length polymorphism (AFLP) analysis (54), and multilocus sequence typing (MLST) (20), have revealed extensive genomic similarities and few consistent differences among isolates currently classified as *B. anthracis*, *B. cereus*, and *B. thuringiensis*. These studies have reinforced the phenotypic argument (15) that the three taxa should be considered a single bacterial species (15, 19).

Despite such biological arguments for unification, a separate species status for these bacteria has been maintained because of their distinctive pathogenic features. Virtually all *B. cereus* group isolates obtained from humans or animals exhibiting the symptoms of anthrax are very closely related to each other, and *B. anthracis* is very likely to be a clone, particularly if associated with the toxin-encoding plasmids pXO1 and pXO2 (29, 30). However, organisms classified as *B. cereus* and *B. thuringiensis* are more diverse, and the evolutionary relationships between all members of the group have yet to be definitively established

* Corresponding author. Mailing address: School of Life Sciences, Heriot Watt University, Edinburgh EH14 4AS, United Kingdom. Phone: 44 (0) 131 451 3464. Fax: 44 (0) 131 451 3009. E-mail: f.g.priest@hw.ac.uk.

† Supplemental material for this article may be found at http://jb.asm.org/.

(21). This is important, not only for understanding the evolution of virulence in the *B. cereus* group, but also for rapidly and accurately characterizing these organisms, a concern which has become of increasing scientific and political importance in recent years.

MLST studies that employ nucleotide sequence analysis to identify genetic variation have been highly successful for characterizing bacterial genetic variation and for developing evolutionary frameworks that interpret this diversity (56). Here we have employed this approach by determining the nucleotide sequences of seven housekeeping gene fragments for 105 representative members of the *B. cereus* group and related organisms. The results demonstrate a largely clonal population structure and indicate that the group comprises at least eight distinct lineages. Two of these lineages centered on *B. anthracis* and *B. thuringiensis* serovar sotto contain strains of a single species, but the remainder are mixed and contain isolates that are currently classified as different species.

MATERIALS AND METHODS

Bacterial isolates. A total of 105 pure cultures, representing *B. anthracis* and related bacteria that were isolated globally over the period 1900–1999, were analyzed (Table 1). The collection comprised isolates that were classified as *B. anthracis* (11 isolates), *B. cereus* (38 isolates), *Bacillus mycoides* (1 isolate), *B. thuringiensis* (53 isolates representing 17 serovars), and *Bacillus weihenstephanensis* (2 isolates). Further details of the strain collection are available at http://pubmlst.org/bcereus/. *B. thuringiensis* isolates were classified on the basis of the presence of a crystal protein and/or insect toxicity, while *B. cereus* isolates lacked a crystal protein. Isolates were named as received and, if necessary, checked for the absence or presence of a crystal protein by light and/or scanning electron microscopy. Bacterial cultures for DNA isolation were grown in nutrient broth (5 ml) containing 0.5% glucose at 30°C until the late exponential phase, which required an incubation period of 16 h for most isolates.

Molecular methods. Chromosomal DNAs were prepared from 1.0-ml aliquots of the cultures by the use of a PureGene DNA isolation kit (Gentra Systems) in accordance with the manufacturer's instructions, with the exception that the lyticase solution was increased to 2.5 μl. Seven genes distributed around the chromosome of *B. anthracis* Ames were chosen for MLST. Four loci, *glpF*, *gmk*, *pta*, and *tpi*, were derived from those used for MLST of *Staphylococcus aureus*, a low-G+C gram-positive bacterium (11). The nucleotide sequences of these loci were obtained from the *B. anthracis* complete genome sequence by BLAST searches and were used to design PCR amplification and nucleotide sequencing oligodeoxyribonucleotide primer sequences. The PCR amplification and nucleotide sequencing primers for the remaining loci, *ilvD*, *pur*, and *pycA*, were designed from the sequences described by Økstad et al. (39). The primers used (annealing temperatures are in parentheses) were as follows: Glp-F, 5′-GCGTTTGTGCTGGTGTAAGT; Glp-R, 5′-CTGCAATCGGAAGGAAGAAG (59°C); Gmk-F, 5′-ATTTAAGTGAGGAAGGGTAGG; Gmk-R, 5′-GCAATGTTCACCAACCACAA (56°C); Gmk2-F (an alternative forward primer for *gmk* that was sometimes necessary), 5′-ATCGTTCTTTCAGGACCTTC (56°C); IlvD-F, 5′-CGGGGCAAACATTAAGAGAA; and IlvD-R, 5′-GGTTCTGGTCGTTTCCATTC (58°C). For emetic strains of *B. cereus*, the following alternative primers were necessary: IlvD2, 5′-AGATCGTATTACTGCTACGG; IlvD2-R, 5′-GTTACCATTTGTGCATAACGC (58°C); Pta-F, 5′-GCAGAGCGTTTAGCAAAAGAA; Pta-R, 5′-TGCAATGCGAGTTGCTTCTA (58°C); Pur-F, 5′-CTGCTGCGAAAAATCACAAA; Pur-R, 5′-CTCACGATTCGCTGCAATAA (56°C); PycA-F, 5′-GCGTTAGGTGGAAACGAAAG; PycA-R, 5′-CGCGTCCAAGTTTATGGAAT (57°C); Tpi-F, 5′-GCCCAGTAGCACTTAGCGAC; and Tpi-R, 5′-CCGAAACCGTCAAGAATGAT (58°C). The same primers were used for DNA sequencing, and the methods used are available at http://pubmlst.org/bcereus/.

Each locus was amplified by PCR and purified by polyethylene glycol precipitation as described previously (10). Nucleotide sequence extension reactions were performed on the purified amplicons by the use of BigDye Ready Reaction mix (ABI Corp), and reaction products were separated and detected on a Prism 3700 or a Prism 310 automated DNA analyzer (ABI Corp.). Nucleotide sequences were determined at least once for each DNA strand and were assembled with the STADEN software package (52). All sequences are available from http://pubmlst.org/bcereus/, while representative sequences have been submitted to GenBank (Table 2). Each unique sequence was assigned an arbitrary allele number by reference to the *B. cereus* group MLST database (http://pubmlst.org/bcereus/), which employed MLSTdbnet software (26). The combination of allele numbers for all seven loci of a given isolate was assigned an arbitrary sequence type (ST); each ST was equivalent to a unique haplotype.

Analysis of sequence diversity. The nucleotide sequences were analyzed with the MEGA (32) and DnaSP (48) packages, which were used to calculate the (uncorrected) *p* distances, mean numbers of nonsynonymous (d_N) and synonymous (d_S) substitutions per site, numbers of differences among various groups of sequences, and numbers of fixed differences and shared polymorphisms among lineages. Distances among concatenated sequences were visualized by split decomposition analysis implemented in the SPLITSTREE program (23), using Hamming distances, which were equivalent to *p* distances.

Phylogenetic analysis. Maximum likelihood (ML) phylogenetic trees were reconstructed by using the general time-reversible model of DNA substitution, with a nucleotide substitution matrix and a shape parameter (α) of a discrete approximation (with four categories) to a gamma distribution of rate heterogeneity among sites, the proportion of invariant sites (I), and the base composition estimated from the empirical data during tree reconstruction. For the ML tree of the concatenated data (see below), these parameter values were as follows: for the general time-reversible substitution model, A→C = 0.60228, A→G = 4.24750, A→T = 0.91374, C→G = 0.32520, C→T = 6.28401, G→T = 1.00000, α = 1.70796, and I = 0.81537. The base compositions were as follows: A = 0.33003, C = 0.16656, G = 0.23004, and T = 0.27307. The parameter values for the individual loci are available upon request. To assess the phylogenetic support for groupings on the tree, we performed a bootstrap resampling analysis (1,000 replications). This analysis was run by using 1,000 replicate neighbor-joining trees estimated by the maximum likelihood substitution model described above. To obtain a general measure of the overall degree of incongruence between trees of each of the seven loci, we compared, for each locus in turn, the likelihood of the ML tree for that locus to those of the ML topologies obtained for the other loci and to 200 randomly generated trees of the same size, with the branch lengths being re-estimated in each case. If the ML trees for each locus were congruent, then all of them would have likelihoods that were higher than those of the random trees (12). All of these analyses were undertaken by using the PAUP* 4.0 software package (53).

RESULTS

Sequence diversity. The MLST gene fragments varied in length from 381 to 504 bp, with average *p* distances of 0.015 (*tpi* gene fragment) to 0.067 (*ilvD* gene fragment). The ratio of nonsynonymous to synonymous mutations (d_N/d_S) was less than one for all loci, from 0.01 (*pur* gene fragment) to 0.110 (*tpi* gene fragment), revealing strong purifying selection in each case. All seven loci examined exhibited base compositions in the range of 38.6 to 44.4 mol% G+C. The most diverse locus in terms of numbers of unique sequences was *glpF*, with 37 MLST alleles, and the least diverse was *gmk*, with 19 MLST alleles (Table 2). The strains were recovered in 59 unique allelic profiles, or STs, which were numbered sequentially, with the exception that the ST-11 designation was not used. Only three STs were represented more than four times in the data set: they were ST-1 (associated with eight *B. anthracis* isolates), ST-8 (associated with three *B. cereus* and five *B. thuringiensis* isolates), and ST-23, which comprised five isolates of *B. thuringiensis* serovar morrisoni. Five STs were present four times in the data set, 3 STs were present three times, 7 STs were present twice, and the remaining 41 STs occurred only once.

Population structure. ML trees were constructed for the single sequence of 2,838 bp of concatenated loci (Fig. 1) and individually for all seven loci (Fig. 2). To test for the presence of similar phylogenetic signals in the eight trees obtained, we performed an ML randomization test by which the similarities in tree topologies among loci were compared to those expected by chance alone. This analysis confirmed that while the topol-

TABLE 1. Strains used for this study and their allocation to lineages

Clade or lineage	ST	Strain	Country of origin	Yr isolated	Original designation[a]	Reference or source[b]
Clade 1 (*B. cereus*)						
Anthracis	ST-1	Ames	United States	1981	*B. anthracis*	
	ST-1	Ames (cured strain)	United States		*B. anthracis*	46
	ST-1	K0610/A0034	China		*B. anthracis*	30
	ST-1	K4834/A0039	Australia	1994	*B. anthracis*	30
	ST-1	K1340/A0062	Poland	1962	*B. anthracis*	30
	ST-1	K1694/A0462	United States	1932	*B. anthracis*	30
	ST-1	K5135/A0463	Pakistan	1978	*B. anthracis*	30
	ST-1	K4596/A0488	United Kingdom	1997	*B. anthracis*	30
	ST-2	K3700/A0267	United States	1937	*B. anthracis*	30
	ST-3	K2478/A0102	Mozambique	1944	*B. anthracis*	30
	ST-3	K2762/A0465	France	1997	*B. anthracis*	30
Cereus I	ST-5	m1545	Brazil	1987	*B. cereus*	MADM
	ST-6	m1564	Brazil	1987	*B. cereus*	MADM
	ST-7	M21	Finland	1998	*B. cereus*	40
	ST-32	ATCC 10987	Canada	1930	*B. cereus*	ATCC
Cereus II (emetic)	ST-26	F4810/72	United States	1972	*B. cereus*	55
	ST-26	S710	United Kingdom	1979	*B. cereus*	41
	ST-26	F3080B/87	United Kingdom	1987	*B. cereus*	40
	ST-26	F3942/87	United Kingdom	1987	*B. cereus*	40
	ST-31	S366	North Sea		*B. cereus*	41
	ST-45	m1293	Brazil	1987	*B. cereus*	MADM
	ST-47	m1576	Brazil	1987	*B. cereus*	MADM
Cereus III	ST-27	F4370/75	United Kingdom	1975	*B. cereus*	41
	ST-57	T10024	Pakistan	1975	*B. thuringiensis* serovar darmstadiensis	IP
	ST-60	T18004	Iraq	1984	*B. thuringiensis* serovar kumamotoensis	IP
Clade 2 (*B. thuringiensis*)						
Kurstaki	ST-8	S57	United States	1975	*B. cereus*	BA
	ST-8	S58	United States	1975	*B. cereus*	41
	ST-8	S59	United States	1975	*B. cereus*	BA
	ST-8	T03a001	France	1961	*B. thuringiensis* serovar kurstaki	IP
	ST-8	T03a075	Iraq	1976	*B. thuringiensis* serovar kurstaki	IP
	ST-8	T03a172	Pakistan	1982	*B. thuringiensis* serovar kurstaki	IP
	ST-8	T03a287	Kenya	1988	*B. thuringiensis* serovar kurstaki	IP
	ST-8	T03a361	Australia	1990	*B. thuringiensis* serovar kurstaki	IP
	ST-13	T04b001	Kenya	1962	*B. thuringiensis* serovar kenyae	IP
	ST-13	T04b054	Iraq	1986	*B. thuringiensis* serovar kenyae	IP
	ST-13	T04b060	Iraq	1987	*B. thuringiensis* serovar kenyae	IP
	ST-13	T04b073	Chile	1993	*B. thuringiensis* serovar kenyae	IP
	ST-15	T07033	Japan	1975	*B. thuringiensis* serovar aizawai	IP
	ST-15	T07058	France	1983	*B. thuringiensis* serovar aizawai	IP
	ST-15	T07180	Spain	1992	*B. thuringiensis* serovar aizawai	IP
	ST-18	T13028	Chile	1993	*B. thuringiensis* serovar pakistani	IP
	ST-25	T05005	United States	1964	*B. thuringiensis* serovar galleriae	IP
	ST-25	T05033	United States	1975	*B. thuringiensis* serovar galleriae	IP
	ST-25	T05144	France	1985	*B. thuringiensis* serovar galleriae	IP
	ST-33	ATCC 10876		1945	*B. cereus*	ATCC
	ST-39	SPS 2		1999	*B. cereus*	40
	ST-40	TSP 11		1999	*B. cereus*	40
	ST-44	m1292	Brazil	1987	*B. cereus*	MADM
	ST-51	T05a015	United States	1977	*B. thuringiensis* serovar canadensis	IP
	ST-54	T07196	Brazil	1993	*B. thuringiensis* serovar aizawai	IP
	ST-59	T18001	Japan	1980	*B. thuringiensis* serovar kumamotoensis	IP
	ST-59	T18002	United States	1980	*B. thuringiensis* serovar kumamotoensis	IP
Sotto	ST-9	NCTC 6474	United Kingdom		*B. cereus*	41
	ST-12	T04002	Canada	1965	*B. thuringiensis* serovar sotto	IP
	ST-12	T04016	Pakistan	1980	*B. thuringiensis* serovar sotto	IP
	ST-12	T04024	Pakistan	1981	*B. thuringiensis* serovar sotto	IP
	ST-12	T15001	United States	1983	*B. thuringiensis* serovar dakota	IP
	ST-16	CCCT 2259	Brazil	1993	*B. thuringiensis* serovar israelensis	27
	ST-16	T08025	France	1988	*B. thuringiensis* serovar morrisoni	IP
	ST-23	T08001	United States	1963	*B. thuringiensis* serovar morrisoni	IP
	ST-23	T08009	United States	1979	*B. thuringiensis* serovar morrisoni	IP
	ST-23	T08012	Pakistan	1980	*B. thuringiensis* serovar morrisoni	IP
	ST-23	T08023	Brazil	1987	*B. thuringiensis* serovar morrisoni	IP
	ST-23	T08031	Brazil	1991	*B. thuringiensis* serovar morrisoni	IP
	ST-49	T04236	Indonsesia	1991	*B. thuringiensis* serovar sotto	IP
	ST-55	T10016	United States	1982	*B. thuringiensis* serovar darmstadiensis	IP
	ST-56	T10003	Germany	1967	*B. thuringiensis* serovar darmstadiensis	IP
	ST-56	T10018	Japan	1982	*B. thuringiensis* serovar darmstadiensis	IP

Continued on following page

TABLE 1—*Continued*

Clade or lineage	ST	Strain	Country of origin	Yr isolated	Original designation[a]	Reference or source[b]
Thuringiensis	ST-10	T01001	Canada	1958	*B. thuringiensis* serovar thuringiensis	IP
	ST-10	T01015	Bulgaria	1962	*B. thuringiensis* serovar thuringiensis	IP
	ST-10	T01022	United States	1964	*B. thuringiensis* serovar thuringiensis	IP
	ST-10	T01326	Chile	1993	*B. thuringiensis* serovar thuringiensis	IP
	ST-20	WSBC 10312	Thailand	1999	*B. cereus*	43
	ST-43	m1278	Brazil	1987	*B. cereus*	MADM
	ST-58	T15006	South Korea	1993	*B. thuringiensis* serovar dakota	IP
Tolworthi	ST-4	ATCC 14579^T	United States	1916	*B. cereus*	ATCC
	ST-14	T06007	Pakistan	1983	*B. thuringiensis* serovar entomocidus	IP
	ST-14	T06010	Pakistan	1983	*B. thuringiensis* serovar entomocidus	IP
	ST-17	T13001	Pakistan	1976	*B. thuringiensis* serovar pakistani	IP
	ST-17	T13004	Pakistan	1980	*B. thuringiensis* serovar pakistani	IP
	ST-19	WSBC 10249	Denmark	1999	*B. cereus*	43
	ST-19	m1280	Brazil	1987	*B. cereus*	MADM
	ST-22	T09010	United States	1979	*B. thuringiensis* serovar tolworthi	IP
	ST-22	T09011	Iraq	1987	*B. thuringiensis* serovar tolworthi	IP
	ST-22	T09024	Indonesia	1991	*B. thuringiensis* serovar tolworthi	IP
	ST-22	T09034	Brazil	1992	*B. thuringiensis* serovar tolworthi	IP
	ST-24	NCIB 6349			*B. cereus*	41
	ST-24	Cal3	Finland	1998	*B. cereus*	40
	ST-24	WSBC 10028	Germany	1999	*B. cereus*	43
	ST-29	S86			*B. cereus*	41
	ST-34	ATCC 11778	United States		*B. cereus*	ATCC
	ST-46	m1550	Brazil	1987	*B. cereus*	MADM
	ST-48	T01246	Iraq	1984	*B. thuringiensis* serovar thuringiensis	IP
	ST-50	T05a001	Canada	1968	*B. thuringiensis* serovar canadensis	IP
	ST-52	T05a019	Pakistan	1980	*B. thuringiensis* serovar canadensis	IP
	ST-53	T07146	Indonesia	1991	*B. thuringiensis* serovar aizawai	IP
Unassigned	ST-28	F4431/3	Indonesia	1973	*B. cereus*	41
	ST-30	S363	North Sea		*B. cereus*	41
	ST-38	ATCC 4342	United States	1900	*B. cereus*	15
Other	ST-21	WSBC 10277	Germany	1999	*B. mycoides*	43
	ST-35	AH621	Norway		*B. cereus*	18
	ST-36	AH647	Norway		*B. cereus*	18
	ST-37	AH684	Norway		*B. cereus*	18
	ST-41	WSBC 10202	Germany	1999	*B. weihenstephanensis*	43
	ST-42	WSBC 10364	Germany	1999	*B. weihenstephanensis*	43

[a] *B. thuringiensis* strains are given serovar designations when they are known.
[b] IP, Collection of *Bacillus thuringiensis* and *Bacillus sphaericus*, Institut Pasteur, Paris, France; BA, Brian Austin, Heriot Watt University, Edinburgh, United Kingdom; and MAMD, Marilena Aquino de Muro, CABI Biosciences, Egham, United Kingdom.

ogies of the eight trees had different likelihoods, indicating that the signals present in these data were not completely congruent and therefore that the population was not entirely clonal, they were far more similar than would be expected by chance alone (Fig. 3). As such, there is a clear signal of phylogenetic history present in these MLST data so that they can be used to reconstruct an evolutionary history of the *B. cereus* group.

The *gmk* tree conformed to the concatenated tree most consistently, and *glpF*, *pycA*, and *tpi* provided strain assignments that were reasonably well correlated with those in the concatenated tree. The *ilvD*, *pta*, and *pur* trees, however, failed to resolve the STs into the major monophyletic groups described below.

Phylogenetic groupings. The overall structure of the ML tree generated from concatenated sequences (Fig. 1) revealed three major phylogenetic groups, with each defined by high bootstrap support values of 85 to 100%. One heterogeneous group based on *B. mycoides* (ST-21) and *B. weihenstephanensis*

TABLE 2. Genetic loci analyzed in this study and their characteristics

Locus	Encoded protein	Genomic position[a]	Fragment length (bp)	Total length of gene (bp)	No. of alleles	Avg *p* distance	d_N/d_S	Representative accession no.
glpF	Glycerol uptake facilitator protein	1014815	381	822	37	0.023	0.108	AY729746–AY729753
gmk	Guanylate kinase (putative)	3688226	504	618	19	0.044	0.022	AY729754–AY729761
ilvD	Dihydroxyacid dehydratase	1736221	393	1674	30	0.067	0.017	AY729762–AY729769
pta	Phosphate acetyltransferase	5122669	414	972	31	0.023	0.019	AY729770–AY729777
pur	Phosphoribosylaminoimidazole carboxamide formyltransferase	306074	348	1,536	30	0.046	0.010	AY729778–AY729785
pycA	Pyruvate carboxylase	3809749	363	3,447	33	0.065	0.028	AY729786–AY729793
tpi	Triosephosphate isomerase	4861379	435	756	34	0.015	0.110	AY729794–AY729801

[a] Based on the *B. anthracis* Ames genome (46).

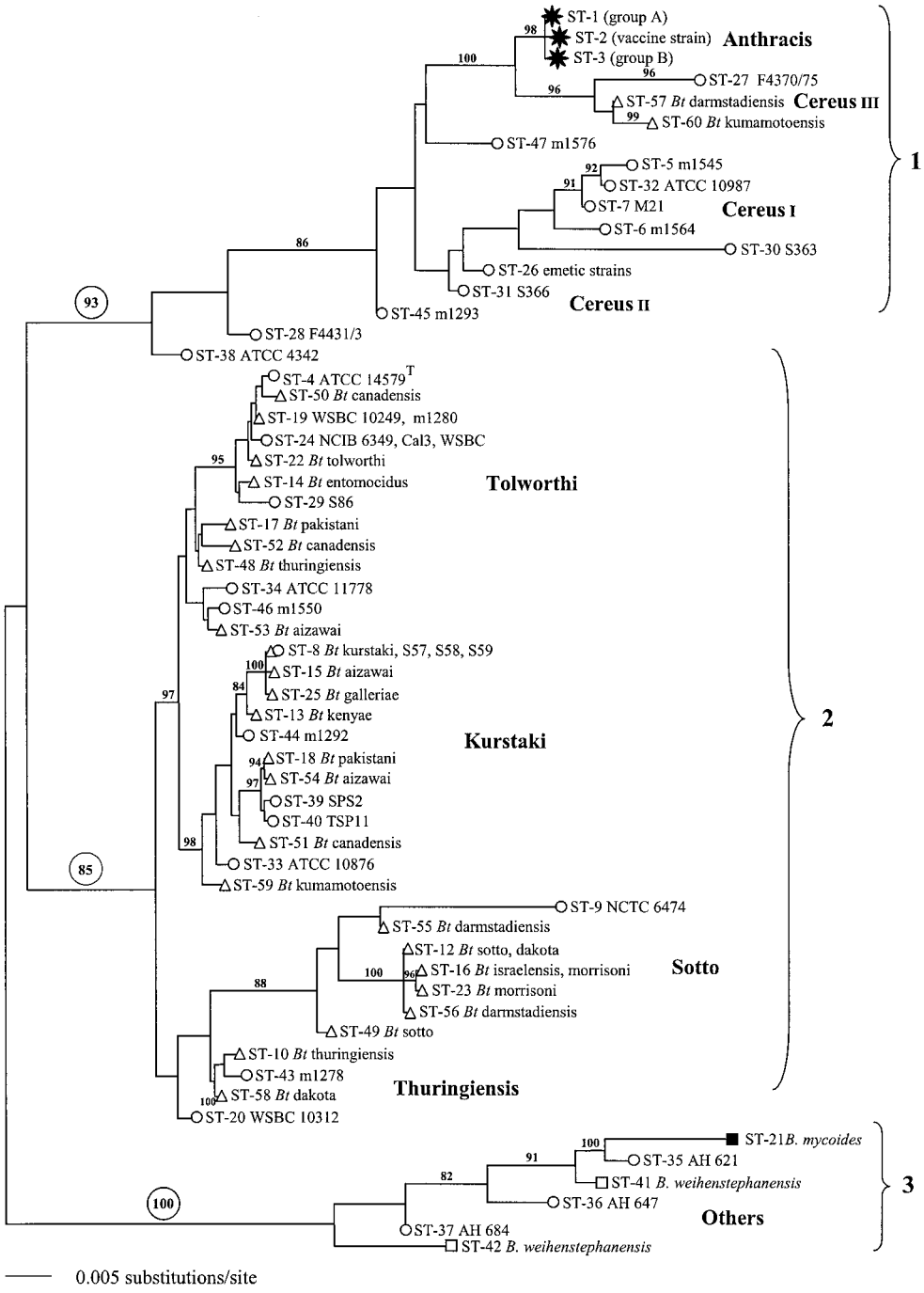


FIG. 1. ML phylogenetic tree for the concatenated gene sequences for the 59 STs included in the study. Strain identifications: ✷, *B. anthracis*; ○, *B. cereus*; △, *B. thuringiensis*; ■, *B. mycoides*; □, *B. weihenstephanensis*. All horizontal branch lengths were drawn to a scale of substitutions per site, and the tree was rooted at the midpoint for the purpose of clarity only. All bootstrap support values of >80% are shown next to the appropriate nodes. The 85% bootstrap value associated with clade 2 excludes the highly divergent ST-9 type.

(ST-41 and ST-42) is referred to here as "others" (Table 1) and was not further considered. A group including *B. anthracis*, numerous *B. cereus* strains, and rare *B. thuringiensis* isolates, notably ST-57 and ST-60, is referred to as clade 1 and labeled *B. cereus* since that was the predominant organism of the cluster (Table 1). Finally, a large cluster that was mostly composed of *B. thuringiensis* strains but that included some *B. cereus* isolates is described as clade 2 and labeled *B. thuringiensis*. The only ambiguous strain in this clade was recovered as ST-9, and the bootstrap value for this clade excluded this highly divergent strain (see below).

The ML tree of concatenated sequences was used as the basis for grouping the allelic profiles into lineages. The validity of these assignments was augmented by examinations of the

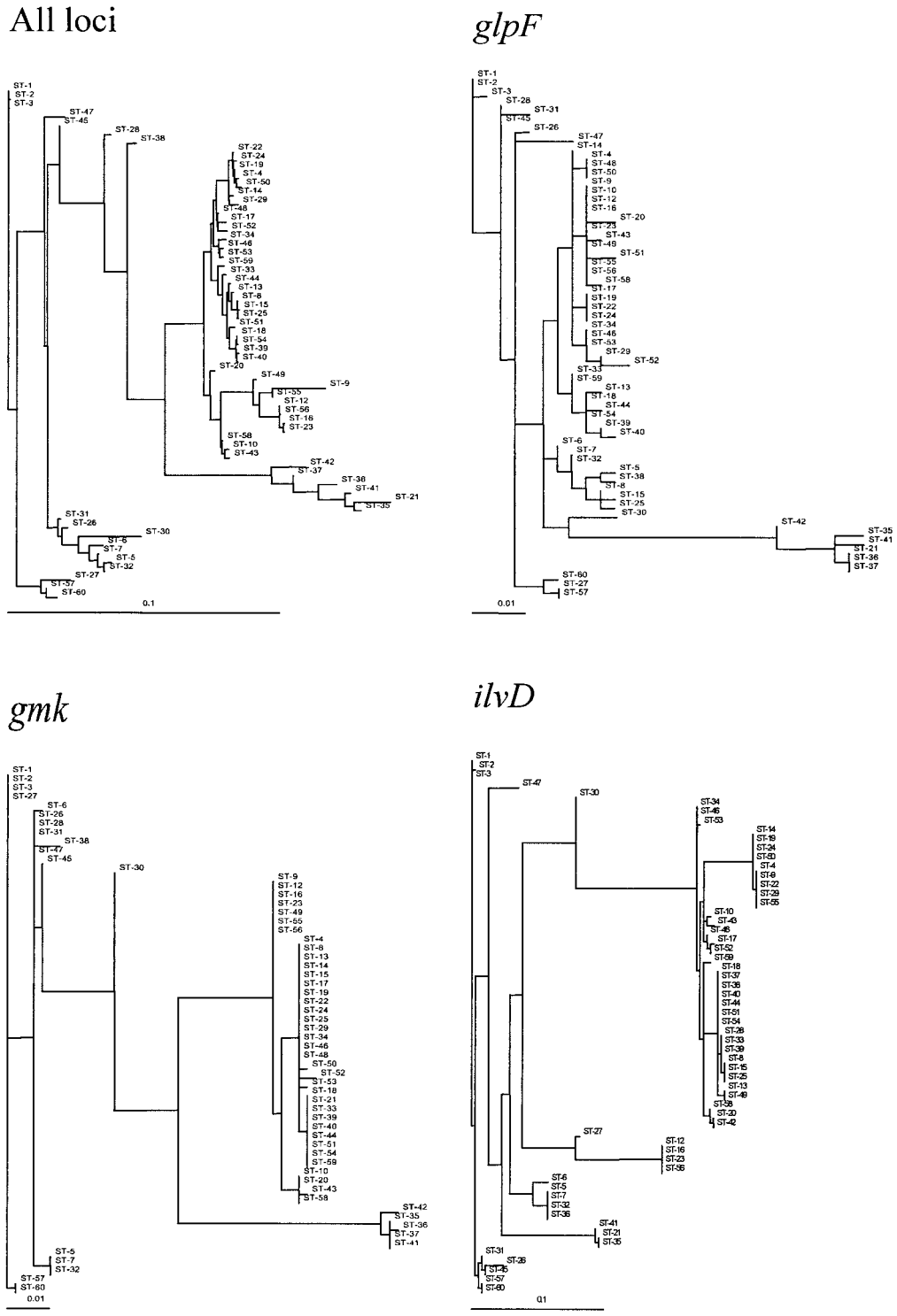


FIG. 2. Maximum likelihood phylogenetic trees obtained for the concatenated sequence and the seven loci. ST designations are given in Table 1. All horizontal branch lengths were drawn to scale.

individual gene trees (Fig. 2) and split decomposition analyses of clades 1 and 2 (see Fig. S1 in the supplemental material). In this way, 50 of the 59 STs were grouped into eight lineages which were assigned names that were as consistent as possible with previous microbiological and serological designations but that were given a unique format (capitalized, nonitalic) to avoid confusion with valid taxonomic labels. The lineage compositions, with the exception of Cereus II, were supported by bootstrap values of >87%. Cereus II was the only lineage for which strain allocation did not correlate with a monophyletic

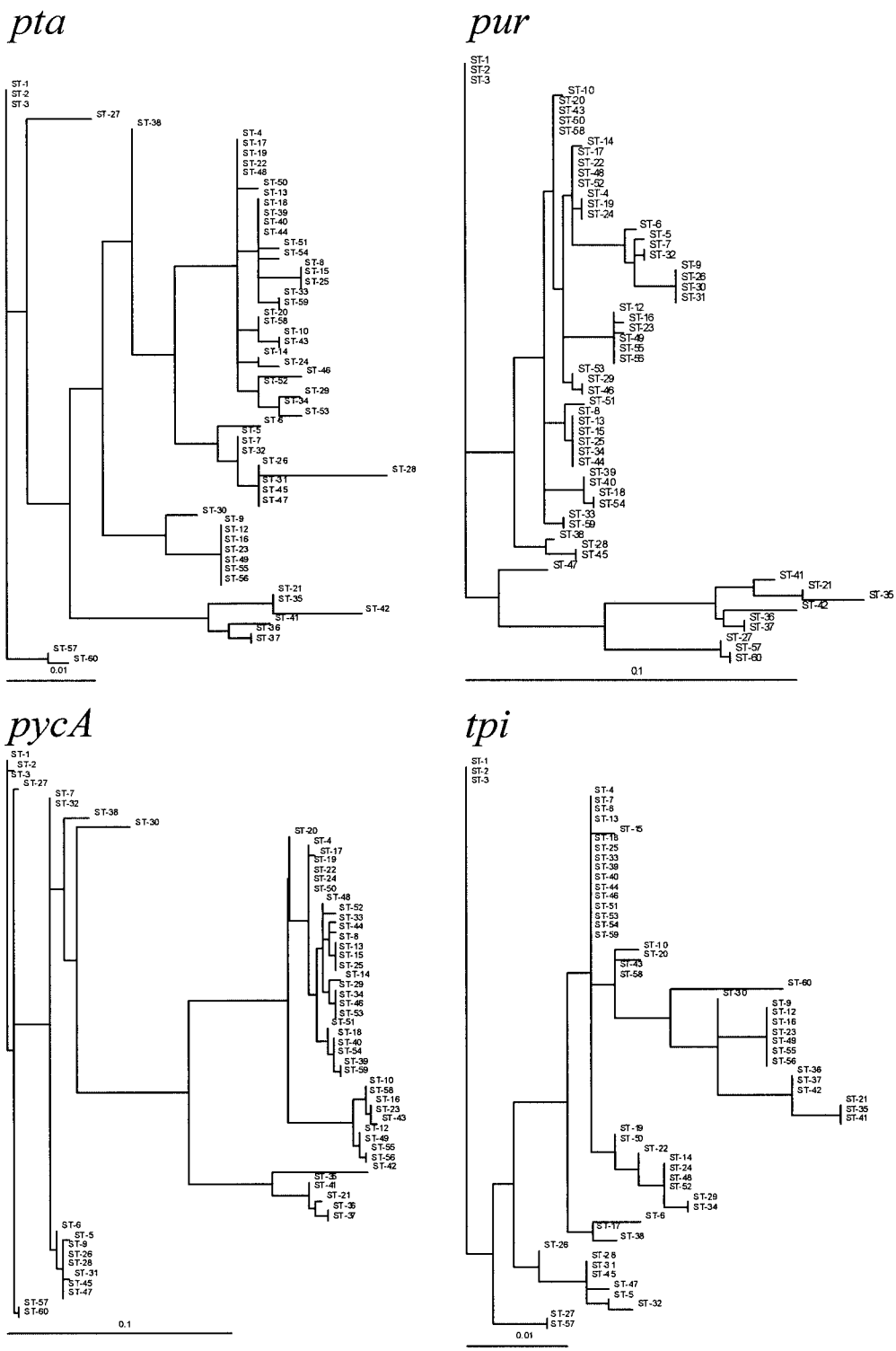


FIG. 2—*Continued.*

group in the ML tree. Its composition was determined by split decomposition analysis, individual alleles that isolates had in common, and the preponderance of fixed differences compared to shared polymorphisms.

Variation among and within lineages. There was an excess of fixed differences compared to shared polymorphisms in pairwise comparisons of all but two of the lineages (Kurstaki and Tolworthi), with the highest numbers of shared polymorphisms occurring between lineages that were more closely positioned in the ML tree of the concatenated sequences (Table 3). Compared with the overall diversity of the data set, there were generally fewer sequence differences within each of the lin-

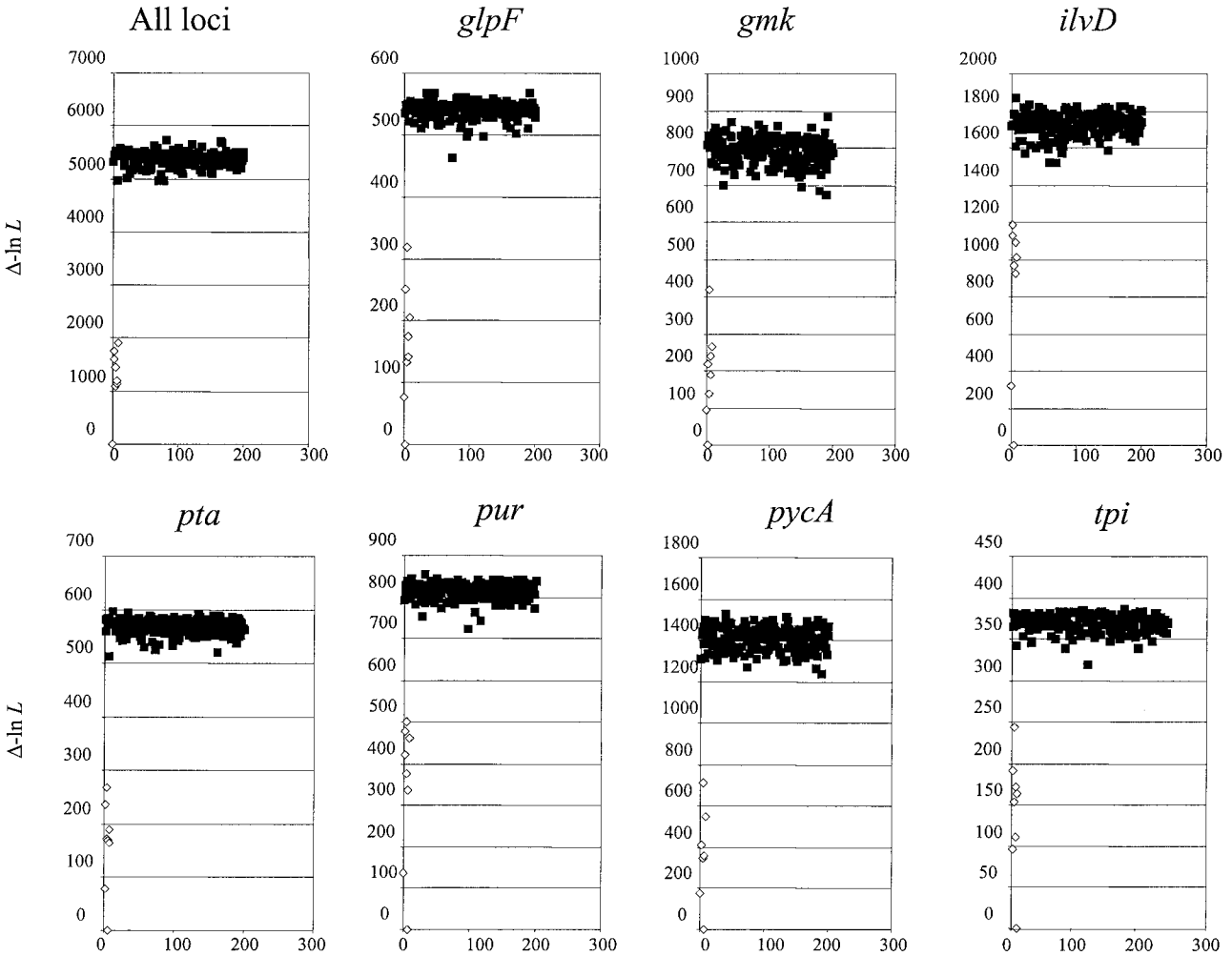


FIG. 3. Maximum likelihood analysis of phylogenetic congruence in the *B. cereus* group. An ML tree that was reconstructed from the data for the concatenated loci and each of the seven loci was compared to each of the eight ML trees, with the branch lengths optimized for each analysis. The differences in likelihood ($\Delta - \ln L$) are shown for each tree (open symbols) and for 200 random trees (closed symbols).

eages at each locus. It was noteworthy that in many cases, multiple sequence differences were due to single genes that were present in individual STs (Table 4), probably as a consequence of lateral gene transfer.

Relationships of lineages to previous designations and phenotypic properties. All of the *B. anthracis* strains exhibited STs that were assigned to the Anthracis lineage (1, 2, and 3); indeed, there were only three polymorphisms identified among the 11 isolates examined, highlighting how closely related these isolates are. Although the two major phylogenetic groups (A and B) of *B. anthracis* (29) were evident as ST-1 and ST-3, respectively, the subdivisions within group A could not be resolved. The vaccine strain used in the United States (V770-NPI-R) was assigned to ST-2, with unique alleles at *ilvD* and

TABLE 3. Numbers of fixed differences across all seven loci and shared polymorphisms among the *B. cereus* group lineages

Lineage	No. of fixed differences (no. of shared polymorphisms)							
	Cereus I	Cereus II	Cereus III	Kurstaki	Tolworthi	Sotto	Thuringiensis	Others
Anthracis	54 (0)	28 (0)	27 (0)	127 (0)	120 (0)	92 (1)	126 (0)	100 (0)
Cereus I		11 (7)	53 (4)	95 (3)	88 (3)	52 (1)	94 (2)	78 (5)
Cereus II			37 (8)	95 (4)	91 (4)	64 (15)	94 (2)	87 (13)
Cereus III				124 (2)	120 (3)	85 (9)	127 (2)	79 (11)
Kurstaki					3 (5)	27 (12)	13 (0)	64 (9)
Tolworthi						24 (16)	11 (6)	67 (10)
Sotto							21 (13)	55 (33)
Thuringiensis								70 (6)
Others								

TABLE 4. Numbers of nucleotide sequence differences within subdivisions of the *B. cereus* group

Subdivision	No. of variable sites in gene							
	All loci	*glpF*	*gmk*	*ilvD*	*pta*	*pur*	*pycA*	*tpi*
All STs	395	50	52	81	50	63	74	25
Clade 1 (*B. cereus*)	158	16	8	44[a]	17	42	14	17
Anthracis	3	1	0	1	0	0	1	0
Cereus I	35	4	3	8	3	4	3	11
Cereus II	54	8	1	17[b]	0	24[c]	1	3
Cereus III	45	2	1	24[d]	0	1	0	8[e]
Clade 2 (*B. thuringiensis*)	195	20	11	49	24	28	49	14
Kurstaki	45	9	1	10	5	12	7	1
Tolworthi	61	6	4	19	9	9	9	5
Sotto	92	0	0	39	0	16[f]	37[g]	0
Thuringiensis	27	4	1	6	1	1	12[h]	2
Others	146	9	33	50	11	22	19	2
Unassigned	114	11	11	29	14	22	17	10

[a] Includes 11 variable sites contributed only from ST-27.
[b] Includes 11 variable sites contributed only from ST-47.
[c] Includes 11 variable sites contributed only from ST-45.
[d] Includes 24 variable sites contributed only from ST-27.
[e] Includes eight variable sites contributed only from ST-60.
[f] Includes 14 variable sites contributed only from ST-9.
[g] Includes 32 variable sites contributed only from ST-9.
[h] Includes 10 variable sites contributed only from ST-20.

pycA, both of which were the result of single nucleotide polymorphisms that do not appear elsewhere in the data set and probably represent mutational changes.

Strains designated *B. cereus* were distributed among several of the lineages, indicating that the characteristics used to identify this species do not necessarily reflect the phylogenetic origins of the strains. For example, several strains of *B. cereus*, including the type strain ATCC 14579, were included in the *B. thuringiensis*-rich clade 2 in lineages Tolworthi, Kurstaki, and Thuringiensis, while most strains were assigned to clade 1. The Cereus I lineage included *B. cereus* ATCC 10987, an atypical xylose-positive strain isolated from cheese, and three other isolates from foods. *B. cereus* strains associated with the emetic form of food poisoning constitute a recognized clone (40) and were recovered here as ST-26 within the Cereus II lineage. The ST-26 strains were isolated from cases of food poisoning, except strain S710, which was isolated from soil. The latter has since been shown to synthesize the emetic toxin cereulide (1.6 ng/ml of culture fluid), consistent with its clonal root with other emetic toxin-forming strains. We included three other STs representing nonemetic *B. cereus* isolates in this lineage, with the lineage being distinguished by a unique *pta5* allele.

The only two strains of *B. thuringiensis* recovered in clade 1 were allocated to the Cereus III lineage together with one strain of *B. cereus* that was isolated from a case of diarrheal food poisoning (Table 1). This lineage was the closest relative of the Anthracis lineage in our collection. Indeed, *B. cereus* F4370/75 was the only strain in the collection that shared an allele with *B. anthracis*, specifically, the *gmk1* allele.

Within clade 2, the Sotto lineage was comprised almost exclusively of *B. thuringiensis* isolates, including both dipteran (serovar israelensis)- and various lepidopteran-active strains. The only *B. cereus* isolate in the Sotto lineage was an outlier (ST-9) which had atypical *pur* and *pycA* alleles, both of which were more commonly associated with the *B. cereus* strains of clade 1, suggesting a chimeric *B. cereus*/*B. thuringiensis* genome. Indeed, the position of ST-9 could not be resolved in the bootstrap analysis. STs in this lineage correlated loosely with previous serovar designations. For example, three strains of *B. thuringiensis* serovar sotto from Pakistan and Canada that were isolated over a 16-year period formed a discrete clone (ST-12), but a strain of *B. thuringiensis* serovar dakota from the United States was also included in this clone, while a fourth strain of serovar sotto was allocated to the unique ST-49 type. Similarly, five strains of *B. thuringiensis* serovar morrisoni were identical (ST-23), but a sixth strain joined a *B. thuringiensis* serovar israelensis isolate in ST-16. Nevertheless, the Sotto lineage was unified by four common alleles, *glp15*, *gmk7*, *pta2*, and *tpi13*, and formed a coherent (with the exception of ST-9) monophyletic group (Fig. 1).

The Kurstaki lineage was the largest in the study, including 12 STs, 8 of which comprised exclusively or predominantly *B. thuringiensis* strains. Several *B. thuringiensis* serovars in this lineage correlated with discrete clones; notably, three strains of serovar aizawai were ST-15, three strains of serovar galleriae were ST-25, and four strains of serovar kenyae were ST-13. ST-8, comprising three strains of *B. cereus* and five strains of *B. thuringiensis* serovar kurstaki, was the only example in the study of *B. cereus* and *B. thuringiensis* strains being allocated to the same ST.

Four strains of *B. thuringiensis* serovar tolworthi from different continents formed the basis of the Tolworthi lineage as ST-22. They were associated with numerous other clones and strains of *B. thuringiensis* representing various serovars and some strains of *B. cereus* (Table 1; Fig. 1). The distinction between lineages Kurstaki and Tolworthi was slight, with only three fixed differences, but the splits graph (see Fig. S1 in the supplemental material) confirmed the divergence that was evident in the ML tree (Fig. 1).

The remaining lineage in clade 2 comprised a clone of four *B. thuringiensis* serovar thuringiensis strains (ST-10) together with a strain of *B. thuringiensis* serovar dakota and two isolates of *B. cereus* (Fig. 1; also see Fig. S1 in the supplemental material).

DISCUSSION

The definition of bacterial species is a continual source of debate (6, 47). While largely pragmatic definitions have been invaluable throughout the history of bacteriology, this has led to a situation in which the degree of genetic diversity seen within different bacterial species varies widely, as does the reproducibility and accuracy of bacteriological identification to the species level. In addition, these definitions may be misleading when a major characteristic used for classification purposes is encoded by a mobile element such as a phage or plasmid. Here we have described a sequence-based multilocus analysis of chromosomally encoded housekeeping genes to explore the relationships among members of the *B. cereus* group of bacteria, which have proved to be particularly refractory to traditional taxonomic investigations. The use of nucleotide sequences has several advantages. The data generated are definitive and reproducible among laboratories, and with the wide availability of complete genome sequences, genetic diversity in any part of the chromosome can be accessed rapidly and

inexpensively. Furthermore, the data can be analyzed by a variety of phylogenetic and population genetic approaches to establish the nature of the variation under examination and to investigate possible evolutionary models for how this variation has arisen.

A preliminary analysis of the variation detected for the seven housekeeping genes used in this study indicated that the genes were similar in their diversity and were all under strong purifying selection. The clonality that is inherent in bacterial populations as a consequence of asexual reproduction can be broken down by recombination, and it is the extent of this lateral gene transfer that sets the degree of clonality in a given bacterial population (51). An analysis of congruence performed on ML trees generated from the concatenated loci and from each of the loci individually indicated that the *B. cereus* group was largely clonal, with evidence for some recombination (Table 4). However, the extent of this recombination is not sufficient to erode the phylogenetic signal in the data, as seen for some other bacterial species (12). Previous estimates of the degrees of association and recombination between alleles (I_A) of strains of the *B. cereus* group similarly concluded that the population structure was clonal with limited recombination (20). With clonal organisms, it is possible to exploit conventional phylogenetic analyses to determine the population structure and evolution as we have done here, although it is necessary to be aware of recombination events because they will compromise the analysis. Since mutation will be more important than recombination in clonal organisms, it is preferable to use nucleotide sequences rather than allelic profiles as the basis for classification and evolutionary analysis of these organisms because allelic profiles do not retain the magnitude of changes between alleles. This contrasts with the case for essentially nonclonal organisms, such as *Neisseria meningitidis*, for which recombination invalidates phylogenetic approaches and for which allelic profiles are a more appropriate basis for such investigations (34).

The phylogenetic tree generated from the concatenated sequences (Fig. 1), together with the individual gene trees, resolved the isolates into eight distinct groups or lineages distributed between two major clades: clade 1 comprises *B. anthracis* and predominantly *B. cereus* strains, and clade 2 comprises largely *B. thuringiensis* strains with sporadic *B. cereus* isolates. A third major clade comprising other species of the *B. cereus* group was also observed. Four of the seven loci (*glpF*, *gmk*, *pycA*, and *tpi*) supported this primary division, while the remaining loci did not assign the STs of clade 1 to a monophyletic group, extending the finding from comparative genome sequences that lateral gene transfer has played a role in metabolic specialization in these bacteria (45). This primary division has been noted in several other population studies of these organisms (20, 21, 44, 57), supporting the contention that the phylogenetic signal is intact despite recent recombinational exchanges, although this will need to be confirmed through the analysis of more loci. In particular, an extensive AFLP analysis of these organisms recognized three major clusters, of which AFLP clusters 1 and 2 map almost perfectly to MLST clades 2 and 1, respectively, while representatives of AFLP cluster 3 were not included in this study (21).

The *B. cereus* group comprises bacteria that have most likely evolved from a saprophyte or insect gut commensal common ancestor, principally by asexual processes. At least eight distinct lineages have arisen, each of which appears to have attained global distribution, although the presence of the unassigned *B. cereus* genotypes represented by ST-28, ST-30, and perhaps ST-38 suggests that more exhaustive sampling would probably identify further lineages and add definition to the extant ones. However, it is notable that *B. cereus* ATCC 4342 (ST-38) was unassigned in both this and a previous MLST study (20) as well as the more extensive AFLP survey (21), suggesting that it is a true atypical strain rather than a representative of a poorly sampled lineage.

The eight lineages correlate closely with the phylogenetic branches of the AFLP analysis described by Hill et al. (21). The mammalian pathogens present in the Anthracis lineage are similar to the insecticidal pathogens in that they form a distinct lineage that has presumably evolved as a consequence of its association with particular plasmids. Despite its wide geographic representation, the Anthracis lineage contains only three sequence genotypes and three polymorphic nucleotides, two of which are only present in a laboratory vaccine strain. It is possible that the latter polymorphisms may indicate further mutational changes in the genome of this strain that could compromise its efficacy as a vaccine. Nevertheless, the high degree of clonality among these strains is consistent with previous reports of the very low genetic diversity of this organism (29, 30), contrasting with the multiple variants observed for the insecticidal lineages. This suggests that the Anthracis lineage is much younger than the insect pathogenic lineages. Hill et al. defined the Anthracis lineage more broadly in their AFLP analysis and included strains of *B. cereus* such as F4431/73 (ST-28), which was unassigned in our study, as well as strains of *B. thuringiensis* in their *B. anthracis* branch F (21). However, in view of the distinctive allelic profiles of *B. anthracis* strains, their strong bootstrap support, and their isolation in a splits graph (see Fig. S1 in the supplemental material), we consider it appropriate to define this lineage more strictly. The Cereus I lineage includes *B. cereus* ATCC 10987, for which there is now a complete genome sequence (45) that confirms its closer phylogenetic affinity to *B. anthracis* than to the *B. cereus* type strain located in clade 2.

Of the lineages in clade 2, the Sotto lineage was comprised almost exclusively of *B. thuringiensis* isolates. This group was recognized as branch A by AFLP analysis and was similarly composed exclusively of *B. thuringiensis* strains, with serovars darmstadiensis, israelensis, morrisoni, and sotto in common between the two studies (21). Serovar assignments did not correlate perfectly with STs in this group (Table 1). However, there is evidence that the ST designation may relate more to insect toxicity than does the serovar. For example, *B. thuringiensis* serovar israelensis (ST-16) constitutes a large, globally widespread clone of highly active, mosquito-pathogenic strains with similar or identical crystal proteins (Cry4Aa, Cry4Ba, Cry10Aa, and Cry11Aa) (1, 27). Interestingly, the only strain of *B. thuringiensis* serovar morrisoni included in ST-16 is also a dipteran pathogen and contains Cry4 toxins (data not shown), making this ST exclusive to Cry4-containing mosquito pathogens. Most *B. thuringiensis* serovar morrisoni strains, on the other hand, are lepidopteran pathogens containing crystals composed of Cry1Aa and Cry1Bc and were assigned to ST-23. The preponderance of crystalliferous bacteria in this lineage is

unique among the four lineages of clade 2. The relatively high numbers of fixed differences and rare shared polymorphisms clearly delineate it from other lineages (Table 3), suggesting that this line of descent represents a particularly successful association between crystal-encoding plasmids and the host genotype.

The Kurstaki lineage corresponds to branch C of the AFLP analysis (21). *B. thuringiensis* serovar assignments in common between the two studies include serovars aizawai, kenyae, kumamotoensis, kurstaki, and galleriae. However, the Tolworthi lineage was not recognized in the AFLP study, and isolates of *B. thuringiensis* serovars canadensis, entomocidus, pakistani, and tolworthi were included with Kurstaki lineage strains in branch C by AFLP (21). The ability to distinguish between lineages Kurstaki and Tolworthi may reflect the higher resolution of MLST, although the few fixed differences and an excess of shared polymorphisms between the two lineages suggest that the division is weak. Nevertheless, there was strong bootstrap support for this division (Fig. 1).

Most serovars in the Kurstaki and Tolworthi lineages were represented by cognate STs, reinforcing the clonal structure of *B. thuringiensis* noted in other studies (13, 14, 42). *B. thuringiensis* serovar kurstaki strains are widespread lepidopteran pathogens containing Cry1 and Cry2 toxins and are often used for biocontrol in agriculture. *B. thuringiensis* serovars aizawai, galleriae, kenyae, and kumamotoensis similarly contain Cry1 toxins, although the exact compositions of the crystals in these bacteria have not been determined (14). This is a group that apparently undergoes extensive sharing of plasmids since all are Cry1-containing types and yet discrete clones are apparent. It seems likely that some purging of diversity gave rise to the extant clones while plasmid promiscuity counters this by enhancing diversity through the generation of novel Cry proteins by recombination (8). The result is a balance represented by numerous clones (ST-8, ST-13, ST-15, and ST-25) distributed among unique STs within the confines of the lineage. ST-8 was the only ST in the study that comprised both *B. thuringiensis* (Cry^+) and *B. cereus* (Cry^-) strains. There was no sign of crystals in sporulated cultures of the *B. cereus* ST-8 strains that were examined by light and electron microscopy, and Western blots using an anti-Cry1 antiserum revealed a trace of crystal protein in strain S58 but none in the other two strains (data not shown). Presumably, these are strains that have lost most or all of the Cry plasmids, supporting the concept of the fluidity of plasmid-borne crystal protein synthesis in this lineage (22).

The Thuringiensis lineage was based on the clone of *B. thuringiensis* serovar thuringiensis isolates from four different countries (ST-10) together with some strains of *B. cereus*. It correlated with branch B defined by AFLP, which also largely comprised *B. thuringiensis* serovar thuringiensis strains (21).

The lineage assignments were confirmed and refined by analyses of the data by split decomposition and examinations of the sequence variation within and among the assigned lineages. On the basis of these results, a redefinition of the nomenclature of the *B. cereus* group was suggested. Each of the eight lineages was considered to be sufficiently distinct to warrant a separate label, and names for these lineages were chosen that were, as far as was possible, consistent with taxonomic designations but distinct in format to avoid confusion with the current system of nomenclature. Such a classification, in which the clone or phylogenetic lineage is given recognition, has been suggested for other clonal taxa, such as the four "species" of the *Mycobacterium tuberculosis* complex, which would become clones Africanum, Bovis, Tuberculosis, and Microti (33). It provides for an effective taxonomy in which, for example, Anthracis can be recognized as a pathogenic lineage but other lineages will contain both entomopathogens (*B. thuringiensis* strains) and nonpathogens (*B. cereus* strains). While it may seem incongruous to retain a separate species status for bacteria that are to be included in a coherent phylogenetic taxon such as a lineage, the implications of renaming *B. thuringiensis* strains as *B. cereus* would be severe for the biocontrol industry. Therefore, for pragmatic reasons, we retained the current species identifications. Nevertheless, clones can be named or coded within lineages and associated where appropriate with specific species and pathogenic traits, such as the emetic clone (ST-26) of the Cereus II lineage or the Morrisoni clone (ST-23) of the Sotto lineage. The MLST scheme described in this study provides the basis for more extensive sampling of the *B. cereus* group such that the population diversity can be more fully estimated and assigned to existing and new lineages and clones in due course.

ACKNOWLEDGMENTS

M.C.J.M. is a Wellcome Trust Senior Research Fellow in Basic Biological Sciences.

We are grateful to various colleagues, in particular M. Aquino de Muro, A.-B. Kolstø, M. Salkinoja-Salonen, and S. Scherer, for the provision of strains and to N. Agata for an estimation of cereulide synthesis by *B. cereus* S710. We thank K. Jolley for assistance with preparation and maintenance of the web site.

REFERENCES

1. **Ankarloo, J., D. A. Caugant, B. M. Hansen, A. Berg, A.-B. Kolstø, and A. Lovgren.** 2000. Genome stability of *Bacillus thuringiensis subsp. israelensis* isolates. Curr. Microbiol. **40:**51–56.
2. **Ash, C., and M. D. Collins.** 1992. Comparative analysis of 23S ribosomal RNA gene sequences of *Bacillus anthracis* and emetic *Bacillus cereus* determined by PCR-direct sequencing. FEMS Microbiol. Lett. **73:**75–80.
3. **Ash, C., J. A. Farrow, M. Dorsch, E. Stackebrandt, and M. D. Collins.** 1991. Comparative analysis of *Bacillus anthracis*, *Bacillus cereus*, and related species on the basis of reverse transcriptase sequencing of 16S rRNA. Int. J. Syst. Bacteriol. **41:**343–346.
4. **Carlson, C. R., D. A. Caugant, and A.-B. Kolstø.** 1994. Genotypic diversity among *Bacillus cereus* and *Bacillus thuringiensis* strains. Appl. Environ. Microbiol. **60:**1719–1725.
5. **Carlson, C. R., T. Johansen, and A.-B. Kolstø.** 1996. The chromosome map of *Bacillus thuringiensis subsp. canadensis* HD224 is highly similar to that of *Bacillus cereus* type strain ATCC 14579. FEMS Microbiol. Lett. **141:**163–167.
6. **Cohan, F. M.** 2002. What are bacterial species? Annu. Rev. Microbiol. **56:** 457–487.
7. **Damgaard, P. H., P. E. Granum, J. Bresciani, M. V. Torregrossa, J. Eilenberg, and L. Vealentino.** 1997. Characterization of *Bacillus thuringiensis* isolated from infections in burn wounds. FEMS Immunol. Med. Microbiol. **18:** 47–53.
8. **de Maagd, R. A., A. Bravo, and N. Crickmore.** 2001. How *Bacillus thuringiensis* has evolved specific toxins to colonize the insect world. Trends Genet. **17:**193–199.
9. **Drobniewski, F. A.** 1993. *Bacillus cereus* and related species. Clin. Microbiol. Rev. **6:**324–338.
10. **Embley, T. M.** 1991. The linear PCR reaction: a simple and robust method for sequencing amplified rRNA genes. Lett. Appl. Microbiol. **13:**171–174.
11. **Enright, M. C., N. P. J. Day, C. E. Davies, S. J. Peacock, and B. G. Spratt.** 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. J. Clin. Microbiol. **38:**1008–1015.
12. **Feil, E., E. C. Holmes, D. E. Bessen, M. S. Chan, N. P. J. Day, M. C. Enright, R. Goldstein, D. W. Hood, A. Kalia, C. E. Moore, J. Zhou, and B. G. Spratt.** 2001. Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. Proc. Natl. Acad. Sci. USA **98:**182–187.

13. **Gaviria Rivera, A., and F. G. Priest.** 2003. Molecular typing of *Bacillus thuringiensis* serovars by RAPD-PCR. Syst. Appl. Microbiol. **26:**254–261.
14. **Gaviria Rivera, A., and F. G. Priest.** 2003. Pulsed field gel electrophoresis of chromosomal DNA reveals a clonal population structure to *Bacillus thuringiensis* that relates in general to crystal protein content. FEMS Microbiol. Lett. **223:**61–66.
15. **Gordon, R. E., W. C. Haynes, and C. H.-N. Pang.** 1973. The genus *Bacillus*. Agriculture handbook no. 427. United States Department of Agriculture, Washington, D.C.
16. **Granum, P. E., and T. Lund.** 1997. *Bacillus cereus* and its food poisoning toxins. FEMS Microbiol. Lett. **157:**223–228.
17. **Guttmann, D. M., and D. J. Ellar.** 2000. Phenotypic and genotypic comparisons of 23 strains from the *Bacillus cereus* complex for a selection of known and putative *B. thuringiensis* virulence factors. FEMS Microbiol. Lett. **188:** 7–13.
18. **Helgason, E., D. A. Caugant, M.-M. Lecadet, Y. Chen, J. Mahillon, A. Lovgren, I. Hegna, K. Kvaloy, and A.-B. Kolstø.** 1998. Genetic diversity of *Bacillus cereus/B. thuringiensis* isolates from natural sources. Curr. Microbiol. **37:**80–87.
19. **Helgason, E., O. A. Økstad, D. A. Caugant, H. A. Johansen, A. Fouet, M. Mock, I. Hegna, and A.-B. Kolstø.** 2000. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*—one species on the basis of genetic evidence. Appl. Environ. Microbiol. **66:**2627–2630.
20. **Helgason, E., N. J. Tourasse, R. Meisal, D. A. Caugant, and A.-B. Kolstø.** 2004. Multilocus sequence typing for bacteria of the *Bacillus cereus* group. Appl. Environ. Microbiol. **70:**191–201.
21. **Hill, K. K., L. O. Ticknor, R. T. Okinaka, M. Asay, H. Blair, K. A. Bliss, M. Laker, P. E. Pardington, A. P. Richardson, M. Tonks, D. J. Beecher, J. D. Kemp, A.-B. Kolstø, A. C. Lee Wong, P. Keim, and M. A. Jackson.** 2004. Fluorescent amplified fragment length polymorphism analysis of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* isolates. Appl. Environ. Microbiol. **70:**1068–1080.
22. **Hu, X., B. M. Hansen, J. Eilenberg, N. B. Hendriksen, L. Smidt, Z. Yuan, and G. B. Jensen.** 2004. Conjugative transfer, stability and expression of a plasmid encoding a *cry1Ac* gene in *Bacillus cereus* group strains. FEMS Microbiol. Lett. **231:**45–52.
23. **Huson, D. H.** 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics **14:**68–73.
24. **Ivanova, N., A. Sorokin, I. Anderson, N. Galleron, B. Candelon, V. Kapatral, A. Bhattacharyya, G. Reznik, N. Mikhailova, A. Lapidus, L. Chung, M. Mazur, E. Goltsman, N. Larsen, M. D'Souza, T. Walunas, Y. Grechkin, G. Pusch, R. Haselkorn, M. Fonstein, S. D. Ehrlich, R. Overbeek, and N. Kyrpides.** 2003. Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. Nature **423:**87–91.
25. **Jackson, S. G., R. B. Goodbrand, R. Ahmed, and S. Kasatiya.** 1995. *Bacillus cereus* and *Bacillus thuringiensis* isolated in gastroenteritis outbreak investigations. Lett. Appl. Microbiol. **21:**103–105.
26. **Jolley, K. A., M.-S. Chan, and M. C. J. Maiden.** 2004. mlstdbNet—distributed multi-locus sequence typing (MLST) databases. BMC Bioinformatics **5:** 86–93.
27. **Kaji, D. A., Y. B. Rosato, V. P. Canhos, and F. G. Priest.** 1994. Characterization by polyacrylamide gel electrophoresis of whole cell proteins of some strains of *Bacillus thuringiensis subsp. israelensis* isolated in Brazil. Syst. Appl. Microbiol. **17:**104–107.
28. **Kaneko, T., R. Nozaki, and K. Aizawa.** 1978. Deoxyribonucleic acid relatedness between *Bacillus anthracis*, *Bacillus cereus* and *Bacillus thuringiensis*. Microbiol. Immunol. **22:**639–641.
29. **Keim, P., A. Kalif, J. M. Schupp, K. K. Hill, S. E. Travis, K. Richmond, D. M. Adair, M. E. Hugh-Jones, C. R. Kuske, and P. Jackson.** 1997. Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymorphism markers. J. Bacteriol. **179:**818–824.
30. **Keim, P., L. B. Price, A. M. Klevytska, K. L. Smith, J. M. Schupp, R. Okinaka, P. J. Jackson, and M. E. Hugh-Jones.** 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. Appl. Environ. Microbiol. **182:**2928–2936.
31. **Kim, W., Y.-P. Hong, J.-H. Yoo, W.-B. Lee, C.-S. Choi, and S.-I. Chung.** 2001. Genetic relationships of *Bacillus anthracis* and closely related species based on variable-number tandem repeat analysis and BOX-PCR genomic fingerprinting. FEMS Microbiol. Lett. **207:**21–27.
32. **Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei.** 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17:**1244–1255.
33. **Lan, R., and P. R. Reeves.** 2001. When does a clone deserve a name? A perspective on bacterial species based on population genetics. Trends Microbiol. **9:**419–424.
34. **Maiden, M. C. J., J. A. Bygraves, J. A. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, and D. A. Caugant.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA **95:** 3140–3145.
35. **Margulis, L., J. Z. Jorgensen, S. Dolan, R. Kolchinsky, F. A. Rainey, and S. C. Lo.** 1998. The Arthromitus stage of *Bacillus cereus*: intestinal symbionts of insects. Proc. Natl. Acad. Sci. USA **95:**1236–1241.
36. **Mock, M., and A. Fouet.** 2001. Anthrax. Annu. Rev. Microbiol. **55:**647–671.
37. **Nakamura, L. K.** 1994. DNA relatedness among *Bacillus thuringiensis* serovars. Int. J. Syst. Bacteriol. **44:**125–129.
38. **Okinaka, R. T., K. Cloud, O. Hampton, A. Hoffmaster, K. Hill, P. Keim, T. Koehler, G. Lamke, S. Kumano, D. Manter, Y. Martinez, D. Ricke, R. Svensson, and P. Jackson.** 1999. Sequence, assembly and analysis of pXO1 and pXO2. J. Appl. Bacteriol. **87:**261–262.
39. **Økstad, O. A., I. Hegna, T. Lindback, A.-L. Rishovd, and A.-B. Kolstø.** 1999. Genome organization is not conserved between *Bacillus cereus* and *Bacillus subtilis*. Microbiology **145:**621–631.
40. **Pirttijarvi, T. S. M., M. A. Andersson, A. C. Scoging, and M. S. Salkinoja-Salonen.** 1999. Evaluation of methods for recognising strains of the *Bacillus cereus* group with food poisoning potential among industrial and environmental contaminants. Syst. Appl. Microbiol. **22:**133–144.
41. **Priest, F. G., M. Goodfellow, and C. Todd.** 1988. A numerical classification of the genus *Bacillus*. J. Gen. Microbiol. **134:**1847–1882.
42. **Priest, F. G., D. A. Kaji, Y. B. Rosato, and V. P. Canhos.** 1994. Characterization of *Bacillus thuringiensis* and related bacteria by ribosomal RNA gene restriction fragment length polymorphisms. Microbiology **140:**1015–1022.
43. **Prüss, B. M., R. Dietrich, B. Nibler, E. Martlbauer, and S. Scherer.** 1999. The hemolytic enterotoxin HBL is broadly distributed among species of the *Bacillus cereus* group. Appl. Environ. Microbiol. **65:**5436–5442.
44. **Radnedge, L., P. G. Agron, K. K. Hill, M. A. Jackson, L. O. Ticknor, P. Keim, and G. L. Andersen.** 2003. Genome differences that distinguish *Bacillus anthracis* from *Bacillus cereus* and *Bacillus thuringiensis*. Appl. Environ. Microbiol. **69:**2755–2764.
45. **Rasko, D. A., J. Ravel, O. A. Økstad, E. Helgason, R. Z. Cer, L. Jiang, K. A. Shores, D. E. Fouts, N. J. Tourrasse, S. V. Angiuoli, J. Kolonay, K. E. Nelson, A.-B. Kolstø, C. M. Fraser, and T. D. Read.** 2004. The genome of *Bacillus cereus* ATCC 10987 reveals metabolic adaptations and a large plasmid related to *B. anthracis* pXO1. Nucleic Acids Res. **32:**977–988.
46. **Read, T. D., S. N. Peterson, N. Tourasse, L. W. Baille, I. T. Paulsen, K. E. Nelson, H. Tettelin, D. E. Fouts, J. A. Eisen, S. R. Gill, E. K. Holtzapple, O. A. Økstad, E. Helgason, J. Rilstone, M. Wu, J. F. Kolonay, M. J. Beanen, R. J. Dodson, L. Brinkac, M. Gwinn, R. T. DeBoy, R. Madpu, S. C. Daugherty, S. C. Durkin, D. H. Haft, W. C. Nelson, J. D. Peterson, M. Pop, H. M. Khouri, D. Radune, J. L. Benton, Y. Mahamoud, L. Jiang, I. R. Hance, J. F. Weidman, K. J. Berry, R. D. Plaut, A. M. Wolf, K. L. Watkins, W. C. Nierman, A. Hazen, R. Cline, C. Redmond, J. E. Thwaite, O. White, S. L. Salzberg, B. Thomason, A. M. Friedlander, P. C. Hanna, A.-B. Kolstø, and C. M. Fraser.** 2003. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. Nature **423:**81–86.
47. **Rossello-Mora, R.** 2003. Opinion: the species problem, can we achieve a universal concept? Syst. Appl. Microbiol. **26:**323–326.
48. **Rozas, J., and R. Rozas.** 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:**174–175.
49. **Schnepf, E., N. Crickmore, J. Van Rie, D. Lereclus, J. Baum, J. Feitelson, D. R. Zeigler, and D. H. Dean.** 1998. *Bacillus thuringiensis* and its pesticidal crystal proteins. Microbiol. Mol. Biol. Rev. **62:**775–806.
50. **Somerville, H. J., and M. L. Jones.** 1972. DNA competition studies within the *Bacillus cereus* group of bacilli. J. Gen. Microbiol. **73:**257–265.
51. **Spratt, B. G., W. P. Hanage, and E. J. Feil.** 2001. The relative contributions of recombination and point mutation to the diversification of bacterial clones. Curr. Opin. Microbiol. **4:**602–606.
52. **Staden, R.** 1996. The Staden sequence analysis package. Mol. Biotechnol. **5:** 233–241.
53. **Swofford, D. L.** 2003. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, Mass.
54. **Ticknor, L. O., A. B. Kolstø, K. K. Hill, P. Keim, M. T. Laker, M. Tonks, and P. J. Jackson.** 2001. Fluorescent amplified fragment length polymorphism analysis of Norwegian *Bacillus cereus* and *Bacillus thuringiensis* soil isolates. Appl. Environ. Microbiol. **67:**4863–4873.
55. **Turnbull, P. C. B., J. M. Kramer, K. Jorgensen, and R. J. Gilbert.** 1979. Properties and production characteristics of vomiting, diarrheal, and necrotizing toxins of *Bacillus cereus*. Am. J. Clin. Nutr. **32:**219–228.
56. **Urwin, R., and M. C. J. Maiden.** 2003. Multi-locus sequence typing: a tool for global epidemiology. Trends Microbiol. **11:**479–487.
57. **Vilas-Boas, G., V. Sanchis, D. Lereclus, M. F. L. Lemos, and D. Bourguet.** 2002. Genetic differentiation between sympatric populations of *Bacillus cereus* and *Bacillus thuringiensis*. Appl. Environ. Microbiol. **68:**1414–1424.